GABRIEL DE ÁVILA OTHERO

A GRAMÁTICA DA FRASE EM PORTUGUÊS

ALGUMAS REFLEXÕES PARA A FORMALIZAÇÃO DA ESTRUTURA FRASAL EM PORTUGUÊS

A GRAMÁTICA DA FRASE EM PORTUGUÊS

ALGUMAS REFLEXÕES PARA A FORMALIZAÇÃO DA ESTRUTURA FRASAL EM PORTUGUÊS

Gabriel de Ávila Othero

A GRAMÁTICA DA FRASE EM PORTUGUÊS

ALGUMAS REFLEXÕES PARA A FORMALIZAÇÃO DA ESTRUTURA FRASAL EM PORTUGUÊS

PORTO ALEGRE
2009

Capa: Vinícius de Almeida Xavier

Diagramação: Gabriela Viale Pereira

Revisão Linguística: Grasielly Hanke Angeli

Dados Internacionais de Catalogação na Publicação (CIP)

O87g Othero, Gabriel de Ávila
A gramática da frase em português [recurso eletrônico] : algumas reflexões para a formalização da estrutura frasal em português / Gabriel de Ávila Othero. – Dados eletrônicos. – Porto Alegre : EDIPUCRS, 2009.
160 p.

Sistema requerido: Adobe Acrobat Reader
Modo de Acesso: World Wide Web: <http://www.pucrs.br/orgaos/edipucrs/>
ISBN 978-85-7430-854-8 (on-line)

1. Português – Gramática. 2. Português – Análise Sintática. 3. Linguística Aplicada. 4. Linguística Computacional. I. Título.

CDD 410.285

Ficha Catalográfica elaborada pelo
Setor de Tratamento da Informação da BC-PUCRS

Av. Ipiranga, 6681 - Prédio 33
Caixa Postal 1429
90619-900 Porto Alegre, RS - BRASIL
Fone/Fax: (51) 3320-3711
E-mail: edipucrs@pucrs.br
http://www.pucrs.br/edipucrs

SOBRE O AUTOR

Gabriel de Ávila Othero é Doutor em Linguística Aplicada pela Pontifícia Universidade Católica do Rio Grande do Sul – PUCRS. O autor tem quatro livros publicados, além de artigos acadêmicos em periódicos nacionais. Também é o editor-fundador da Revista Virtual de Estudos da Linguagem – ReVEL.

AGRADECIMENTOS

Este livro eletrônico passou pela leitura crítica de diversos colegas e professores. Alguns leram o texto integralmente; outros leram alguns capítulos. Todos, porém, colaboraram para minha redação final do livro. Deixo aqui registrado meu agradecimento às seguintes pessoas: Ana T. Ibaños, Carlos Prolo, João Paulo Cyrino, Jorge Campos Costa, Leda Bisol, Magdiel M. Aragão Neto, Mário A. Perini, Mathias Schaff Filho e Sérgio de Moura Menuzzi.

Os erros e inconsistências que persistem são de minha inteira responsabilidade.

LISTA DE FIGURAS, QUADROS E TABELAS

LISTA DE SÍMBOLOS E ABREVIATURAS

∅ = elemento vazio

Adj = adjetivo

Adv = advérbio

AdvP = sintagma adverbial

AP = sintagma adjetival

CFG = *Context-free grammar*

CP = *complementizer phrase*

DCG = *Definite Clause Grammars*

Det = determinante

DP = *determiner phrase*

GB = Teoria da Regência e Ligação

HPSG = *Head-driven Phrase Structure Grammar*

IP = *inflectional phrase*

LFG = *Lexical Functional Grammar*

N = nome (ou substantivo)

NP = sintagma nominal

NumP = sintagma numeral

P = preposição

PB = português brasileiro

PLN = Processamento de Linguagem Natural

PossP = sintagma possessivo

PP = sintagma preposicional

PSG = *Phrase Structure Grammar*

QP = *quantifier phrase*

S = sentença

V = verbo

VP = sintagma verbal

Para quem gosta de certezas e seguranças, tenho más notícias: a gramática não está pronta. Para quem gosta de desafios, tenho boas notícias: a gramática não está pronta. Um mundo de questões e problemas continua sem solução, à espera de novas idéias, novas teorias, novas análises, novas cabeças.

Mário A. Perini

SUMÁRIO

PRIMEIRÍSSIMAS PALAVRAS

Uma professora de Semântica disse uma vez algo mais ou menos assim à nossa turma: "estudar ciência e fazer Linguística é, antes de tudo, revelar problemas. Vocês devem se dar por satisfeitos se saírem de nosso curso com mais dúvidas do que tinham quando chegaram. Aqueles de vocês que acham que vieram aqui apenas para buscar respostas ficarão decepcionados. Vocês verão que sairão do curso com mais dúvidas e questionamentos do que quando chegaram. E aí então, eu ficarei satisfeita".

Este livro tem também esse caráter. O que as próximas páginas irão mostrar é o resultado de um trabalho cuidadoso, modesto, honesto e atento que reflete minhas dúvidas enquanto estudioso da gramática da língua portuguesa. Meu trabalho irá trazer alguns questionamentos meus que tentei solucionar, alguns problemas já conhecidos por gramáticos e linguistas e alguns desafios que os linguistas computacionais por vezes enfrentam ao trabalhar com a sintaxe das línguas naturais e, especialmente, da língua portuguesa. Muitos dos problemas ficarão sem solução, seja porque eu não tive a capacidade de resolvê-los, seja porque permaneci fiel demais à teoria que adotei para trabalhar com os dados, mesmo quando a teoria se mostrava limitada frente aos fatos da língua.

De qualquer forma, espero que este trabalho possa servir também para guiar novas dúvidas e novas discussões sobre novos e velhos desafios no estudo da sintaxe da língua portuguesa. Sigo Mário Perini quando afirma que "para um lingüista não é vergonha dizer 'não sei'" (Perini, 2006: 186) e assim dou por iniciado este livro.

1 UMAS PRIMEIRAS PALAVRAS

Rem tene, verba sequentur[1].

Marcus Porcius Cato

Ray Jackendoff conta uma piada já consagrada entre os linguistas computacionais: sempre que os informatas resolvem pedir auxílio aos linguistas teóricos em algum de seus projetos de Processamento de Linguagem Natural (PLN), seus programas acabam se tornando menos eficientes[2]. Isso porque, como veremos, esses dois tipos de pesquisadores nem sempre têm os mesmos objetivos no estudo da linguagem humana. Aristóteles já havia afirmado que "o objetivo do conhecimento teórico é a verdade, enquanto o do conhecimento prático é a eficácia"[3]. Utilizando-nos de uma comparação, podemos nos arriscar a dizer que o linguista tem um compromisso com a verdade, e o informata, com a eficácia[4].

As análises que iremos desenvolver aqui tentarão olhar para os dois lados da Avenida Linguística Computacional. Ou seja, por um lado, iremos nos preocupar em descrever a estrutura interna da sentença em português brasileiro com base em um formalismo linguístico consagrado nos estudos sintáticos (a saber, o formalismo da *Phrase Structure Grammar*, nos moldes propostos pela teoria X-barra). Por outro lado, estaremos também pensando em uma possível e potencial aplicação de nossas análises em aplicativos que envolvam o processamento sintático computacional. Nosso estudo certamente será uma matéria da Sintaxe, mas estaremos sempre "de olho" em uma possível implementação computacional. Iremos propor uma descrição sintática da sentença em português (na verdade, de alguns tipos de estruturas sentenciais) com a pretensão de auxiliar não apenas o linguista, mas também o informata e o linguista computacional que estiverem interessados na compreensão da estrutura sintática interna da sentença e dos sintagmas em PB. Partiremos de análises já consagradas na literatura sobre a sintaxe do português (especialmente Pontes, 1973; Lobato, 1975, 1986; Perini, 1989, 2000; Lemle, 1984; Luft, 1986; Raposo, 1992; e Mioto et al., 2004), além de análises formais voltadas à aplicação

[1] "Domine o assunto, e as palavras irão fluir".
[2] Jackendoff (2002: 267): "Computational linguists, I am told, joke that every time they hire a theoretical linguist, their programs become less effective".
[3] Aristóteles (1969 [350 a.C.]), livro II, parte 1.
[4] Perseguimos essa ideia com algum aprofundamento em Othero (2008).

computacional, como Othero (2004 e 2006), Pagani (2004) e David (2007), e apresentaremos nossas análises, que tentarão ser mais precisas, abrangentes e coerentes do que as análises que estudamos.

Como dissemos, argumentaremos que nossas regras gramaticais poderão ser interessantes tanto ao linguista computacional quanto ao sintaticista formal. Isso porque elas serão elaboradas, por um lado, pensando na implementação computacional (mostraremos um exemplo de implementação computacional de regras sintagmáticas do português no terceiro capítulo) e, por outro lado, elas serão sintaticamente consistentes e seguirão um modelo largamente utilizado em teoria gramatical, o formalismo da teoria X-barra.

***

Há uma pergunta que todo o pesquisador deve saber responder sobre seu próprio trabalho (e tentaremos respondê-la ao longo deste livro, mas especialmente aqui neste capítulo): qual é a relevância do seu trabalho? Ou seja, qual é a sua importância?

Eis aí uma boa pergunta, porém ambígua: a pergunta pode se referir a nosso livro, especificamente, ou à área em que o livro se insere. Em outras palavras, a pergunta pode ser entendida como "por que nosso texto é relevante? Qual é a sua contribuição original?" ou como "por que a descrição e o processamento sintático das línguas naturais – e do português, em específico – são relevantes?".

Já que temos tempo e espaço não nos falta, nos arriscaremos[5] a responder às duas perguntas. Debrucemo-nos primeiramente sobre a última: qual é a importância de se estudar a língua de maneira formal, matematizada (por que não?), pensando em sua implementação computacional? E mais especificamente: qual é a importância de se estudar sobre a descrição e o processamento sintático de uma língua?

O processamento sintático é uma área de interface entre a Sintaxe e a Linguística Computacional. Esta última é a área da ciência Linguística preocupada com o tratamento computacional da linguagem e das línguas naturais. Ela envolve diversos campos de estudo da Linguística Teórica e Aplicada, e seus primeiros desenvolvimentos datam dos anos 1950. A área teve grande impulso principalmente graças a esforços para o desenvolvimento de programas de tradução automática nas décadas de 1950 e 1960 e está

[5] Decidimos conscientemente evitar o uso da mesóclise em nosso texto.

intrinsecamente ligada a desenvolvimentos na área da Inteligência Artificial[6].

De acordo com Grisham (1992: 1),

> o potencial para o processamento de linguagem natural foi reconhecido bem cedo no desenvolvimento de computadores, e trabalhos em Lingüística Computacional – basicamente para tradução automática – começaram na década de 1950 em diversos centros de pesquisa. O rápido crescimento na área, no entanto, aconteceu principalmente a partir do final dos anos 1970.[7]

E a Sintaxe tem se mostrado uma área da Linguística de extremo valor para os desenvolvimentos em Linguística Computacional. O nível de formalização sintática alcançado pelas pesquisas em Linguística tem proporcionado bons resultados não somente para pesquisas de cunho teórico envolvendo o estudo de línguas naturais, mas também para o desenvolvimento de aplicativos computacionais, como *parsers, taggers*, corretores ortográficos e gramaticais, tradutores automáticos, etc.[8] Para Sag, Wasow & Bender (2003: 14), "uma das áreas mais promissoras para a aplicação da teoria sintática é no desenvolvimento de tecnologias robustas que lidam com a linguagem natural"[9].

O estudo da sintaxe de uma língua pode ser essencial para seu tratamento computacional em diversos outros níveis de descrição linguística, e normalmente trabalhos de implementação computacional em uma língua envolvem o tratamento sintático ou morfossintático dessa língua. Para Sag, Wasow & Bender (2003: 14),

> até mesmo quando tecnologias usáveis que lidem com a linguagem natural podem ser desenvolvidas sem se utilizarem da pesquisa em gramática, normalmente o que acontece é que elas podem se tornar mais robustas ao incluírem um componente de base sintática.[10]

---

[6] Cf. Sparck Jones (1994), Russel & Norvig (1995) e Hutchins (2000).

[7] Trecho original: "The potential for natural language processing was recognized quite early in the development of computers, and work in computational linguistics – primarily for machine translation – began in the 1950s at a number of research centers. The rapid growth in the field, however, has taken place mostly since the late 1970s".

[8] Cf. Nijholt & Sikkel (1993), Bick (1996) e Garside, Leech and McEnery (1997).

[9] Trecho original: "one of the most promising areas for applying syntactic research is in the development of useful and robust natural language technologies".

[10] Trecho original: "even where usable natural language technologies can be developed that are not informed by grammatical research, it is often the case that they can be made more robust by including a principled syntactic component". Porém, o processamento sintático não é essencial. Wong & Wu (1999), por exemplo, defendem a ideia de que o desenvolvimento de *parsers* robustos não precisa sequer contar com gramáticas e regras sintáticas: "Construir uma gramática que defina/descreva precisamente uma língua natural é muito difícil, se não impossível. Uma gramática construída à mão ou induzida automaticamente costuma ser incompleta e generalizadora demais. (...) Uma gramática explícita [para efetuar o *parsing* de uma língua] não é necessária. (...) [Acreditamos na] possibilidade de melhorar a precisão e a velocidade de *parsing* se eliminarmos uma gramática explícita" (p. 121-2). Ver também, a esse respeito, Kac & Manaster-Ramer (1986), que admitem que "em Lingüística Computacional, existe uma forte (se não universal) dependência em gramáticas sintagmáticas como meio de se representar a estrutura sintática de uma língua" (p. 156); eles

O fato de se estudar a sintaxe de uma língua antes de partir para outros tipos de estudos em sua formalização também pode estar relacionado à **sintaticização**, como chama Culicover (1997). Ou seja, costumamos pensar que o tratamento sintático pode preceder outros estudos formais da linguagem. Para Culicover (1997: 12), a sintaticização é

> a tática de pegar qualquer fenômeno que esteja plausivelmente ligado à estrutura sintática e caracterizá-lo em termos sintáticos. A sintaticização é, no mínimo, uma maneira de tornar materiais lingüísticos complexos passíveis de análise e de produzir hipóteses e previsões claras.[11]

De qualquer forma, concordamos com o que dizem Sag, Wasow & Bender (2003: 16):

> Desenvolver tecnologias realmente robustas que lidem com a linguagem natural – ou seja, programas que irão te permitir interagir com o computador na TUA língua e não na língua do computador – requer análises detalhadas e cuidadosas sobre a estrutura gramatical da língua e como ela influencia o significado. Atalhos que se baseiem em heurísticas, tentativas de adivinhação ou simples usos de *templates* irão inevitavelmente levar a erros.[12]

Para Bateman, Forrest & Willis (1997: 166),

> um dos principais objetivos da área de PLN nos últimos dez anos tem sido produzir um "analisador gramatical", ou ***parser***, de **abrangência ampla**. Para muitos aplicativos de PLN, o desafio é produzir um *parser* que poderá ser capaz de analisar automática e estruturalmente de maneira correta, de acordo com um esquema de *parsing* definido, qualquer sentença do inglês que possa ocorrer naturalmente, **sem restrições**, de uma gama de gêneros textuais tão vasta quanto possível.[13] (grifos dos autores).

Doze anos já se passaram desde que essas linhas de Bateman, Forrest & Willis

---

pretendem mostrar que as gramáticas sintagmáticas e as regras gramaticais de reescrita são desnecessárias e redundantes para as representações linguísticas de que um *parser* necessita.

[11] Trecho original: "Syntacticization is the tactic of taking every phenomenon that is plausibly tied to syntactic structure and characterizing it in syntactic terms. At the very least, Syntacticization is a way of rendering complex linguistic material amenable to analysis and of producing clear hypotheses and predictions."

[12] Trecho original: "Building truly robust natural language technologies – that is, software that will allow you to interact with your computer in YOUR language, rather than in ITS language – requires careful and detailed analysis of grammatical structure and how it influences meaning. Shortcuts that rely on semantic heuristics, guesses, or simple pattern-matching will inevitably make mistakes".

[13] Trecho original: "one of the major aims of NLP over the past ten years has been to produce a **wide-range** 'grammatical analyser' or **parser**. For many NLP applications, the challenge is to produce a parser which will automatically be able to structurally analyse correctly, according to a defined *parsing* scheme, any sentence of naturally occurring **unrestricted** English, from as wide a range of genres as possible".

foram escritas, e diversos *parsers* já foram desenvolvidos ao longo desse tempo. Nenhum deles, porém, foi ainda capaz de alcançar o objetivo proposto pelos autores. Por isso, esse ainda continua sendo um dos principais objetivos na agenda da área de PLN, de maneira geral, e do processamento sintático das línguas naturais, de maneira específica.

Sabemos que os estudos formais da linguagem fizeram grandes avanços no século XX, mas não há ainda regras suficientemente **matematizadas**, formais, explícitas e **gerativas** que possibilitem o desenvolvimento de um *parser* que, de certa forma, "compreenda" alguma língua natural. O trabalho de descrição sintática do português que apresentamos aqui certamente pretende revelar alguns *insights* sobre a estrutura sintática das sentenças em PB e formalizar outros *insights* que alguns estudos sintáticos pioneiros já revelaram.

Por que será que ainda não existem descrições linguísticas tão acuradas, formais e matematizadas que nos permitam desenvolver *parsers* que respondam ao desafio proposto por Bateman, Forrest & Willis que vimos acima?

Um pouco de história pode nos auxiliar com a compreensão do problema. Uma das maiores decepções dos linguistas e dos cientistas da computação das décadas de 1950 e 1960 foi justamente esta: ao tentar fazer com que um computador "compreendesse" a linguagem humana, passou-se a perceber que ela é de difícil formalização. McDonald & Yazdani (1990: 6) têm uma frase muito interessante: "o grande problema em PLN é (...) que nós ainda não compreendemos completamente o funcionamento da linguagem humana"[14]. A frase é interessante especialmente para o linguista, cujo trabalho é justamente **compreender (completamente?) o funcionamento da linguagem humana**.

De acordo com Newmeyer (1980: 2),

> houve um sentimento generalizado entre os lingüistas nos anos 1950 de que os problemas fundamentais da análise lingüística haviam sido solucionados e tudo o que restava era preencher alguns detalhes. As afirmações teórico-metodológicas básicas do trabalho de Bloch, "A Set of Postulates for Phonemic Analysis" (1948), e do trabalho de Harris, *Methods in Structural Linguistics* (1951), pareciam tornar qualquer trabalho teórico mais básico algo desnecessário. Na verdade, **muitos lingüistas achavam que os procedimentos tinham sido trabalhados de maneira tão detalhada, que computadores poderiam assumir o trabalho de análise lingüística. Tudo o que se precisava fazer (em teoria) seria colocar os dados no computador e uma gramática sairia prontinha!**

[14] Trecho original: "The major problem in NLP, however, is that we still do not fully understand the human natural language process".

> Também se acreditava que os computadores poderiam resolver outro problema lingüístico tradicional – a tradução. A idéia de uma máquina tradutora foi primeiramente sugerida apenas em 1949 (em um memorando de Warren Weaver). Em 1955, tal trabalho de tradução estava acontecendo em três países, em meia dúzia de instituições. Esses seis anos foram o suficiente para converter os céticos (...).[15] (grifos nossos)

Mesmo um trabalho que parecia simples *a priori*, como a tradução automática de sentenças de uma língua para outra, mostrou-se um empreendimento extremamente complexo, dada a riqueza da linguagem. Este trecho de Baker (2001: 3) também é bastante ilustrativo e nos parece interessante:

> Nós geralmente achamos que o xadrez é uma atividade intelectual requintada que pode ser dominada apenas pelos melhores e mais brilhantes. Qualquer pessoa comum, no entanto, pode falar um horror em um vernáculo entendível sem necessariamente ser considerada inteligente. Ainda assim, mesmo que existam programas de computador que agora podem derrotar os melhores jogadores de xadrez do mundo, não existe nenhum sistema artificial que possa se igualar a um falante médio de cinco anos de idade no que diz respeito a falar e compreender sua língua materna.[16]

De fato, já temos campeões de xadrez que são computadores[17]. No entanto, ainda não temos nenhum programa de computador que consiga "compreender" ou "produzir" linguagem tão bem como um ser humano[18]. No entanto, os estudos formais da linguagem têm contribuído para o avanço da Linguística Computacional. Alguns estudos em sintaxe formal, especialmente de cunho gerativo, já avançaram muito no conhecimento do

---

[15] Trecho original: "There was a widespread feeling among linguists in the 1950s that the fundamental problems of linguistic analysis had been solved and that all that was left was to fill in the details. The basic-theoretical-methodological statements of Bloch's "A Set of Postulates for Phonemic Analysis" (1948) and Harris's *Methods in Structural Linguistics* (1951) seemed to render any more basic theoretical work unnecessary. In fact, many linguists felt that the procedures had been so well worked out that computers could take over the drudgery of linguistic analysis. All one would have to do (in principle) would be to punch the data into the computer and out would come the grammar."
There was also a feeling that computers could solve another traditional linguistic problem – translation. The idea of machine translation had been first suggested (in a memorandum by Warren Weaver) only in 1949. By 1955, such translation work was going on in three countries at half a dozen institutions. These six years were enough to convert the skeptics (…)".

[16] Tradução adaptada. Cf. trecho original: "We usually think of chess as a quintessentially intellectual activity that can be mastered only by the best and brightest. Any ordinary person, in contrast, can talk your ear off in understandable English without necessarily being regarded as intelligent for doing so. Yet although computer programs can now beat the best chess players in the world, no artificial system exists that can match an average five-year-old at speaking and understanding English".

[17] Baker provavelmente está se referindo ao Deep Blue, uma máquina de jogar xadrez produzida pela IBM, que conseguiu vencer o então campeão mundial Gary Kasparov, em 1997.

[18] Há uma espécie de competição em que vários programas de conversação, conhecidos como *chatterbots*, tentam se passar por seres humanos. É o Prêmio Loebner, que foi baseado a partir de ideias propostas pelo teste de Turing (cf. Turing, 1950, Hodges, 1983, Oppy & Dowe, 2005).

funcionamento da linguagem humana, e muitos linguistas computacionais têm se beneficiado desses avanços. Para Newmeyer (1980: 250), "aprendemos mais sobre a natureza da linguagem nos últimos 25 anos do que nos 2.500 anos anteriores"[19]. Carnie (2002: 5), por exemplo, afirma que

> a tese subjacente da gramática gerativa é a de que as sentenças são geradas por um conjunto subconsciente de procedimentos (como programas de computador). Esses procedimentos são parte de nossas mentes (...). O objetivo da teoria sintática é modelar esses procedimentos.[20] (grifos nossos).

Acreditamos que o ser humano tenha um conhecimento inato do funcionamento sintático de sua língua. Há regras a que os falantes obedecem ao produzirem os enunciados, regras "subconscientes" que os falantes não sabem explicitar. O trabalho em sintaxe formal e computacional pode ser entendido desta maneira então: deve-se explicitar a um computador as regras de funcionamento sintático da língua para que o computador possa então "compreender" a linguagem. Um programa como um *parser*, que apresentaremos no capítulo 3, analisa uma sentença em linguagem natural (o português, por exemplo), dizendo se a sentença é gramatical ou agramatical na língua. Além disso, ele atribui às sentenças gramaticais a sua correta estrutura sintática (de acordo com a gramática que foi implementada no programa). Para isso, é necessário o trabalho prévio do linguista: é ele quem deve formalizar as regras sintáticas de uma língua de tal maneira que elas possam ser implementadas e tratadas computacionalmente. Então, uma estratégia para o desenvolvimento de um *parser* seria investi-lo do conhecimento já explicitado sobre os processos de formação das sentenças e dos sintagmas da língua. Ou seja: deveríamos ensinar ao *parser* tudo aquilo que as gramáticas sabem. Evidentemente, seria preciso antes formalizar as regras gramaticais para que elas pudessem ser manipuladas pelo computador. E aí entra este nosso trabalho: tentaremos propor regras formais de descrição sintática do português que deem conta da análise da estrutura dos sintagmas e de uma vasta gama de sentenças em PB.

---

[19] Trecho original: "(…) more has been learned about the nature of language in the last 25 years than in the previous 2500". Essa parece ser uma ideia mais ou menos generalizada entre os linguistas gerativos. Para Strozer (1994: ix), os estudos linguísticos "avançaram mais nos últimos dez ou quinze anos do que nos 30 anos anteriores". Jairo Nunes (em comunicação pessoal) disse que frases como essas são comuns em seções introdutórias de textos recentes dos estudos gerativos. Cf. também a esse respeito Harris (1993), Chomsky (1988, 1995, 2002) e Thomas (2004).

[20] Trecho original: "The underlying thesis of generative grammar is that sentences are generated by a subconscious set of procedures (like computer programs). These procedures are part of our minds (...). The goal of syntactic theory is to model these procedures".

Certamente nossas regras não serão capazes de descrever *todas* as estruturas dos sintagmas em português. Teremos algum mérito se as regras que apresentarmos conseguirem prever apenas estruturas bem formadas e dar conta de uma ampla variedade de sintagmas. Veremos que essa é uma tarefa complicada já em sua natureza. Mas contamos com o auxílio de estudos pioneiros da sintaxe do português para propor novas regras categoriais de reescrita que consigam analisar corretamente algumas estruturas da sentença em PB. Ao tentar propor uma gramática do português com regras de reescrita, lembramos uma passagem de Archangeli & Langendoen (1997: viii), que comparam a tarefa do linguista com a tarefa de um pescador:

> Os lingüistas são, então, confrontados com dois problemas relacionados. Um é garantir que a gramática de uma língua particular seja capaz de abranger todas as expressões que podem pertencer à língua. A outra é garantir que a gramática seja capaz de distinguir aquelas expressões que pertencem à língua daquelas que não pertencem à língua.
> O problema pode ser comparado ao problema de um pescador tentando pegar em sua rede todos os peixes de um determinado tipo em uma certa região, mas nada mais (nenhum peixe de outra espécie, nenhuma outra criatura que não seja peixe, etc.). A rede ideal deverá ser grande e fina o suficiente para pegar todos os peixes desejados pelo pescador (os desejáveis), e deverá ser projetada para permitir que os peixes indesejados e outras criaturas (os indesejáveis) possam escapar. Mas pode não ser possível criar este tipo de rede. Qualquer rede que seja grande e fina o suficiente para apanhar todos os desejáveis pode precisar também apanhar alguns indesejáveis[21].

Esperamos que o leitor lembre-se do problema e da complexidade da tarefa de descrição gramatical ao chegar ao capítulo 4, quando apresentaremos nossas regras categoriais de reescrita para a descrição dos sintagmas e da sentença em PB. No próximo capítulo, apresentaremos o modelo sintático que iremos utilizar em nossas descrições do português, a teoria X-barra.

---

[21] Trecho original: "Linguists are thus faced with two related problems. One is to ensure that the grammar of a particular language is able to encompass all of the expressions that can reasonably be supposed to belong to that language. The other is to ensure that the grammar is able to distinguish those expressions which belong to the language from those which do not.
The problem can be compared to that of a fisherman trying to catch in a net all the fish of certain types in a certain area, but nothing else (no other types of fish, no other creatures, etc.). The ideal net would be large and fine enough to gather all the desired fish (the desirables), and be designed to allow the undesired fish and other creatures (the undesirables) to escape. But it may not be possible to construct such a net. Any net which is large and fine enough to catch all the desirables may of necessity also catch some undesirables".

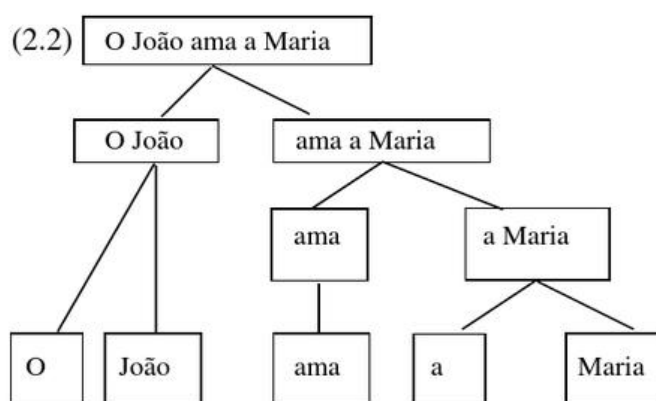
2 O FORMALISMO GRAMATICAL

Fornecer definições formais de noções importantes faz com que seja mais fácil achar inconsistências e insuficiências na teoria. Em teorias descritas informalmente é difícil achar problemas. (...) É praticamente impossível determinar as conseqüências de teorias informais e, assim, de mostrá-las falsas. Teorias formalizadas, por sua vez, levam à descoberta de inconsistências e a soluções para problemas outros, além daqueles para os quais a teoria foi originalmente desenhada para tratar. Quando usamos uma linguagem formal, somos forçados a ser mais precisos; todas as noções usadas em uma definição devem ser definidas elas mesmas.

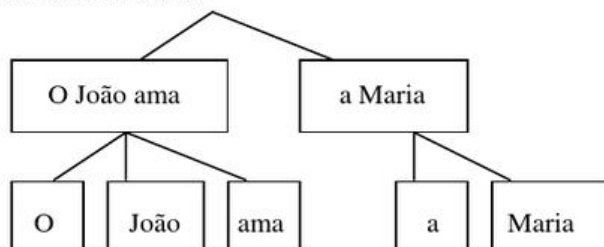
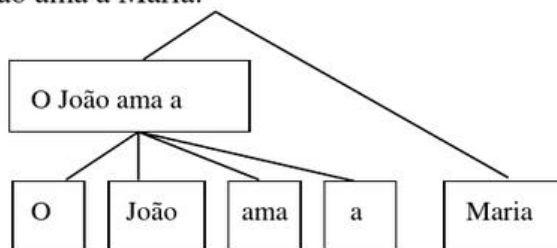
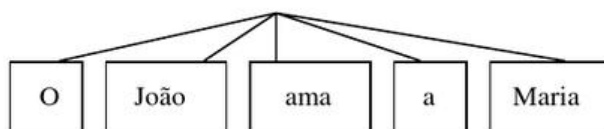
Marcello Modesto

2.1 A GRAMÁTICA SINTAGMÁTICA

Ao efetuarmos análises das sentenças em português, estaremos pressupondo que elas apresentem uma determinada organização sintática **estrutural**. Por exemplo, ao analisarmos (2.1), iremos fazer algo como (2.2):

(2.1) O João ama a Maria.

Não analisamos a sentença (2.1) sem nenhum critério, agrupando quaisquer grupos de palavras aleatoriamente. Se assim fosse, poderíamos tê-la analisado de várias outras maneiras, formando qualquer tipo de agrupamento interno na frase, até chegarmos aos itens lexicais, como em (2.3) e (2.4). Poderíamos até mesmo supor que a sentença não

tenha qualquer organização interna, sendo formada simplesmente por um aglomerado linear de palavras, como em (2.5):

(2.3) O João ama a Maria.

(2.4) O João ama a Maria.

(2.5) O João ama a Maria.

Não iremos encarar a sentença como um mero aglomerado de palavras, unidas uma a outra de qualquer forma. Há entre o nível da palavra e o da frase uma outra forma de organização, que é o sintagma (ou constituinte). De acordo com Radford (1981: 69),

> um certo agrupamento de palavras é um constituinte somente se tiver uma ou mais das seguintes propriedades:
> (i) Comporta-se distribucionalmente como uma única unidade estrutural, isto é, é recorrente como uma unidade única em uma variedade de outras posições nas frases
> (ii) Pode ser coordenado com outro agrupamento similar
> (iii) Não permite intrusão interna de elementos parentéticos (a intrusão geralmente sendo permitida apenas nas fronteiras de constituintes maiores, especialmente sintagmas)
> (iv) Pode ser substituído por (ou servir de antecedente para) uma pro-forma
> (v) Pode ser omitido, sob condições de discurso apropriadas[22.]

[22] Trecho original: "A given string of elements is a constituent just in case it has one or more of the following properties:
(i) It behaves distributionally as a single structural unit – i. e. it recurs as a single unit in a variety of other

Isso quer dizer que podemos realizar diferentes "testes" para identificarmos um constituinte, como os testes de coordenação, interpolação, anáfora, entre outros[23]. Esses testes permitirão identificar os limites dos constituintes, confirmando ou não nossa própria intuição linguística. Essa maneira de "enxergar" e analisar a sentença e seus constituintes é conhecida como *Phrase Structure Grammar*, ou PSG.

A PSG é um formalismo de base gerativa referido por Chomsky (especialmente em 1955, 1956 e 1957) que, nos anos 1980, teve um grande impulso graças a estudos sintáticos formais não transformacionais bastante interessantes, como a *Generalized Phrase Structure Grammar* (GPSG) e, mais tarde, a *Head-driven Phrase Structure Grammar* (HPSG)[24]. Uma PSG possui quatro componentes centrais: (i) um conjunto finito de símbolos terminais; (ii) um conjunto finito de símbolos não terminais; (iii) um conjunto finito de regras gramaticais; e (iv) um símbolo inicial.

Em regras como

(2.6) S → A B

(2.7) A → a

(2.8) B → b

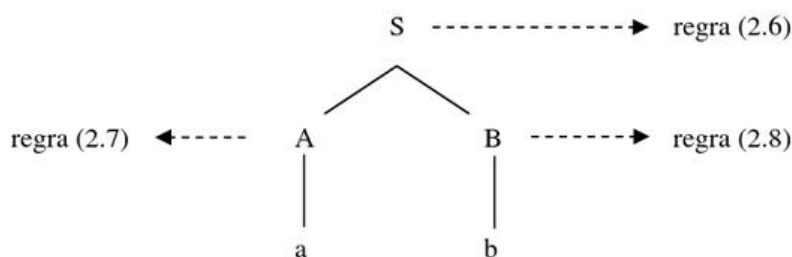
temos um símbolo inicial **S**, símbolos não terminais **A** e **B** e símbolos terminais **a** e **b**. Há diferentes maneiras de se entender as regras acima[25]; entenderemos as regras de reescrita como instruções para um *parser top-down*, ou seja, um mecanismo de construção de estruturas que começa seu funcionamento de cima para baixo. De acordo com a regra (2.6), **S** pode ser expandido em **A** e **B**; seguindo a regra (2.7), **A** pode ser expandido em **a**; finalmente, a regra (2.8) diz que **B** pode ser expandido em **b**. Como resultado, obtemos a seguinte estrutura sintagmática:

---

sentence positions
(ii) It can be coordinated with another similar string
(iii) It does not readily permit intrusion of parenthetical elements internally (intrusion generally being permitted only at the boundaries of major – especially phrasal – constituents)
(iv) It can be replaced by, or serve as the antecedent of, a proform
(v) It can be omitted, under appropriate discourse conditions"

[23] Ver a esse respeito Radford (1981, 1988), Aarts & Haegeman (2006) e Perini (2006).

[24] Sobre a GPSG, cf. Gazdar et al. (1985); sobre a HPSG, cf. Pollard & Sag (1987, 1994).

[25] Cf. Borsley (1999), cap. 2.

Há dois pontos importantes que devemos notar: (i) as regras de uma PSG são **gerativas**, ou seja, são explicitamente descritas; e (ii) elas apresentam um símbolo não terminal à esquerda e um símbolo (ou uma sequência de símbolos) terminal à direta. Aqui estamos nos referindo a uma gramática PSG **livre de contexto** (**CFG**, do inglês *context-free grammar*). Uma PSG livre de contexto é

> um tipo específico de sistema formal que tem se mostrado bastante útil em precisar a caracterização de línguas computacionais e que também serve como ponto inicial para muitos trabalhos em teoria sintática. As CFGs consistem em um símbolo inicial, um léxico finito com palavras classificadas em categorias gramaticais e um conjunto finito de regras da forma A → ω, onde A é um símbolo único (representando um tipo de sintagma) e ω é uma seqüência finita de categorias lexicais e/ou sintagmáticas.[26] (Sag, Wasow & Bender, 2003: 558)

Ainda de acordo com Sag, Wasow & Bender (2003: 26), as CFGs são "o ponto inicial para a maioria das tentativas sérias de se desenvolver gramáticas formais para as línguas naturais". Soames & Perlmutter (1979: 26) vão mais longe e afirmam que "as regras livres de contexto são o único tipo de regras de estrutura sintagmática de que se precisa nas gramáticas das línguas naturais" [27].

Todas as regras de boa formação da sentença e dos sintagmas em PB que apresentaremos seguirão o estilo de regras de uma gramática de estrutura sintagmática livre de contexto. Na verdade, utilizaremos uma "instância" da PSG, a saber, o esquema proposto pela **teoria X-barra**.

---

[26] Trecho original: "A context-free grammar is a particular type of formal system that has proved very useful in the precise characterization of computer languages and also serves as the starting point for much work in syntactic theory. CFGs consist of an initial symbol, a finite lexicon with words classified into grammatical categories, and a finite collection of rules of the form A → ω, where A is a single symbol (representing a type of phrase), and ω is a finite string of lexical and/or phrasal category".

[27] Trechos originais: "(...) the starting point for most serious attempts to develop formal grammars for natural languages" (Sag, Wasow & Bender, 2003: 26); "(...) context-free rules are the only kind of phrase structure rules needed in the grammar of natural languages" (Soames & Perlmutter, 1979: 26). Ver, no entanto, Chomsky (1956), Pullum & Gazdar (1982), Shieber (1985), Savitch et al. (1987), Wong & Wu (1999) e Hopcroft et al. (2001), para críticas às regras de PSGs livres de contexto e sua utilização no trabalho de descrição linguística. Uma boa introdução ao assunto pode ser vista em Gazdar (1983).

2.2 TEORIA X-BARRA

A teoria X-barra é o modelo que iremos utilizar para dar conta da organização interna dos sintagmas e da sentença. De acordo com Alencar (no prelo),

> a teoria X-barra é um dos pilares do modelo Princípios e Parâmetros da lingüística gerativa. Trata-se de uma concepção restritiva da gramática de estrutura sintagmática, a qual permite análises psicolingüisticamente mais realistas e computacionalmente mais elegantes do que as praticadas anteriormente no âmbito da gramática gerativa. Desse modo, essa teoria tem sido empregada também na lingüística computacional, em análises baseadas em formalismos gramaticais de natureza não transformacional que operam com a unificação de traços.

No entanto, a teoria X-barra não é restrita ao estudo sintático em P&P e GB. A teoria X-barra é, na verdade, uma teoria sobre a estrutura das sentenças e dos constituintes nas línguas naturais. Ela é uma teoria adotada por diferentes modelos de descrição linguística. Como apontam Kornai & Pullum (1990: 2), a teoria X-barra é

> discutida em quase todos os livros modernos de sintaxe, e é rotineiramente aceita como uma teoria da estrutura sintagmática em uma variedade de escolas distintas de pensamento gramatical, como a Teoria da Regência e da Ligação (GB), *Lexical-Functional Grammar* (LFG) e *Generalized Phrase Structure Grammar* (GPSG).[28]

Um dos motivos pelos quais resolvemos adotar o modelo da teoria X-barra para nossas descrições da sintaxe do PB foi justamente este: a teoria X-barra é uma teoria sintática coerente que já está no mercado há mais de trinta anos[29] e tem rendido estudos interessantes nas línguas naturais. Além do mais, conhecemos apenas uma proposta abrangente de descrição sintática da estrutura da sentença do português seguindo o modelo, a proposta de Lemle (1984)[30]. Aqui, em nosso texto, pretendemos rever as

[28] Trecho original: "(…) is discussed in almost all modern textbooks of syntax, and it is routinely assumed as a theory of phrase structure in a variety of otherwise widely differing schools of grammatical thought such as government-binding theory (GB), lexical-functional grammar (LFG), and generalized phrase structure grammar (GPSG)". Além disso, o modelo proposto pela teoria X-barra também é adotado em trabalhos em **Teoria da Otimidade** (cf. Legendre, 2001, Kuhn, 2003, entre outros), é parcialmente utilizado pela proposta do ***Simpler Syntax*** (cf. Culicover & Jackendoff, 2005, cf. especialmente p. 22) e pelo formalismo da **HPSG** (cf. Pollard & Sag, 1994, e Sag, Wasow & Bender, 2003).

[29] As bases da teoria X-barra remetem aos trabalhos de Chomsky (1970), Jackendoff (1977) e Stowell (1981).

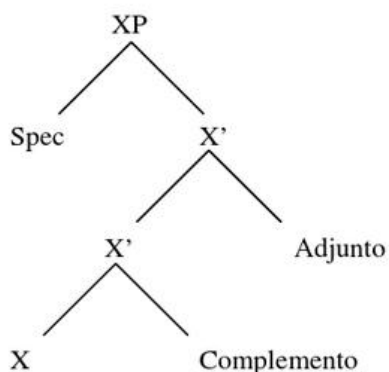
[30] Evidentemente, há outros pesquisadores que utilizam a teoria X-barra para analisar fragmentos do português, ou introduzir questões de análise do português, como Lobato (1986), Raposo (1992), Costa (1998), Mioto et al. (2004), etc.

descrições de Lemle e atualizá-las ao modelo corrente de análise sintática no modelo da X-barra. Além disso, existem trabalhos de descrição sintática do português bastante interessantes (como Pontes, 1973; Lobato, 1975, 1986; Luft, 1986; Perini, 1989, 2000) que não foram devidamente formalizados e não se utilizam do modelo teórico da X-barra. Talvez a formalização desses estudos seguindo o modelo da teoria X-barra possa revelar aspectos interessantes, tanto para o modelo como para a descrição sintática do português.

Por fim, acreditamos que a formalização de tais descrições do português no modelo X-barra poderá nos fornecer uma gramática passível de implementação computacional. Sabemos que a teoria ainda não foi devidamente trabalhada em termos computacionais e ainda pode revelar potencial. Há alguns trabalhos interessantes para a língua inglesa (Chen & Wu, 1988; Black, 1997; Fong, 1999, 2001; Fouskakis, 2005; entre outros) e também para a língua portuguesa (Pagani, 2004; David, 2007; Alencar, no prelo), incluindo nossos próprios trabalhos, como Othero (2004, 2006), Menuzzi & Othero (2008) e Othero & David (no prelo). Esperamos dar uma contribuição para os estudos gramaticais em língua portuguesa ao propor nossa descrição da sentença e de seus constituintes seguindo o molde da X-barra.

Para Mioto et al. (2004: 46), a teoria X-barra é

> o módulo da gramática que permite representar um constituinte. Ela é necessária para explicitar a natureza do constituinte, as relações que se estabelecem dentro dele e o modo como os constituintes se hierarquizam para formar a sentença.

Em outras palavras, a teoria X-barra trata sobre a organização sintática **estrutural** dos constituintes da sentença. Ela apresenta um "molde", um **modelo** para a descrição sintática dos tipos de sintagmas nas línguas naturais. Para nossas análises, partiremos do seguinte molde para a descrição da estrutura arbórea dos constituintes:

Onde temos a projeção máxima de um sintagma **XP** (ou **X''**), a projeção intermediária **X'** e seu núcleo **X** (ou $\mathbf{X^0}$). Há ainda três outras posições: **Spec**, **Adjunto** e **Complemento**. O complemento está mais próximo ao núcleo (é **irmão** de X), enquanto o adjunto vem ao lado do constituinte intermediário (é **irmão** de X')[31]. Repare que essa é uma definição **estrutural** de complementos e adjuntos (ela não leva em consideração detalhes de qualquer outra natureza que não esteja refletida na estrutura da sentença – como as propriedades semânticas dos constituintes, por exemplo). Essa estrutura arbórea permite refletir o fato de que os complementos costumam estar mais próximos ao núcleo do que os adjuntos[32]. A posição de **Spec** será ocupada pelo sujeito da sentença ou por algum outro elemento modificador externo.

A representação dos constituintes em estrutura X-barra traz algumas vantagens em relação a uma simples PSG. Como vimos no esquema acima, o modelo X-barra prevê corretamente uma organização hierárquica que seja maior do que o X e menor do que o XP, além de fazer uma distinção sintática entre complementos e adjuntos[33].

Assumiremos o esquema da teoria X-barra que apresentamos acima, juntamente com alguns princípios da teoria[34]:

a) Endocentricidade
b) Lexicalidade
c) Sucessão
d) Uniformidade
e) Maximalidade

[31] Essa é **uma** maneira de trabalhar com o modelo da teoria X-barra, seguindo, por exemplo, Lobato (1986), Haegeman (1995), Culicover (1997), Radford (1997a), Carnie (2002), Santorini & Kroch (2007), entre outros. Mas ela não é consensual. Há autores, como Chomsky (1986b), Mioto et al. (2004) e Cançado (2005), que tratam os adjuntos como filhos da projeção máxima XP e não da projeção intermediária X'.

[32] A bibliografia que argumenta em favor dessa ideia é vasta. Sérgio Menuzzi (em comunicação pessoal), diz que representar um complemento próximo de seu núcleo na estrutura arbórea sugere que as relações de dominância sintática refletem iconicamente o grau de "intimidade semântica" do modificador em relação ao núcleo. Em línguas VO, como o português, os adjuntos costumam vir em posição final (VO+Adjuntos). Em línguas OV, como o alemão, os adjuntos deixam sua posição final e passam a ocupar a posição anterior aos complementos (Adjuntos+OV). Isso revela uma generalização interessante: os núcleos verbais e seus argumentos internos parecem formar um constituinte coeso, deixando elementos modificadores ocuparem uma posição satélite.

[33] Motivações empíricas para se assumir esses dois pontos são apresentadas largamente em Radford (1988) e Raposo (1992).

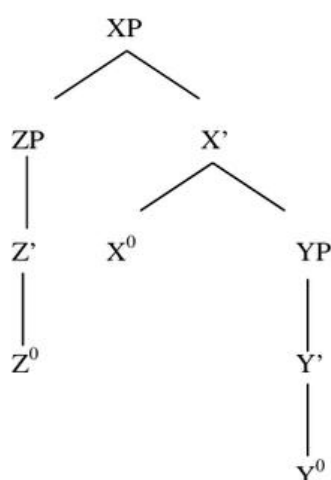
[34] Iremos nos basear especialmente nos princípios discutidos no artigo clássico de Kornai & Pullum (1990). Deixemos claro que esses princípios não são aceitos unanimemente; cada um deles já foi desafiado – cf., por exemplo, o próprio texto de Kornai & Pullum (1990), além de Kayne (1984, 1994) e Culicover & Jackendoff (2005).

f) Opcionalidade

g) Binaridade

A **endocentricidade** é o princípio que garante que cada constituinte **XP** tenha o núcleo **X** (e não um núcleo **Y** ou **Z**):

Todos os sintagmas são encabeçados por um núcleo (uma projeção zero, X^0), de mesma natureza que a projeção máxima.

O princípio da **lexicalidade** postula que cada elemento não terminal seja a projeção de um elemento terminal, como mostra o diagrama abaixo:

Juntamente com o princípio da endocentricidade, garantimos que cada elemento XP ou X' tenha seu núcleo X^0. Esse não foi um princípio respeitado em Lemle (1984), Lobato (1986) e Othero (2004), por exemplo, que seguiram propostas mais antigas da teoria X-barra, que ainda contavam com um elemento não terminal **S**, da sentença, que não possuía qualquer núcleo. Esse elemento se ramificava em dois outros elementos não terminais, NP e VP, por exemplo, não apresentando, portanto, um núcleo terminal, ou lexical, $\mathbf{X^0}$, o que acabou desrespeitando também o princípio da endocentricidade

(voltaremos ao assunto em seguida).

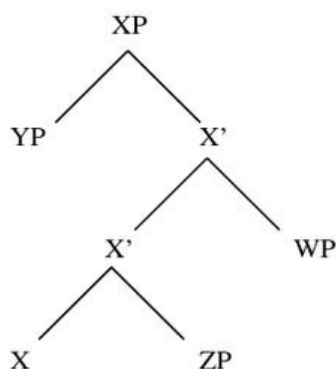
O princípio de **sucessão** diz respeito à relação de dominância entre os elementos constituintes do sintagma. De acordo com esse princípio, cada X^{n+1} domina X^n, onde $n \geq 0$. Na verdade, iremos respeitar apenas parcialmente o princípio de sucessão, uma vez que iremos propor (seguindo Lobato, 1986; Mioto et al., 2004; e Raposo, 1992, entre outros) que um elemento X^n pode dominar um outro elemento X^n (isso acontecerá quando tivermos sintagmas verbais com mais de um complemento, ou com mais de um modificador adverbial, por exemplo). Adotaremos o esquema

$X^n \rightarrow \ldots X^m \ldots$ (onde $m = n$ ou $n\text{-}1$).

O princípio da **uniformidade** garante que todas as projeções máximas tenham o mesmo número de barras. Em nossas análises, o valor de **n** máximo de barras será igual a 2 (que é o número mais usualmente adotado para o número máximo de barras em uma projeção máxima, cf. Radford, 1981, 1988; Lobato, 1986; Raposo, 1992, entre outros):

Apesar de haver mais ou menos um consenso de que o número máximo de barras seja igual a 2, este também não é um ponto pacífico entre os sintaticistas. No texto clássico de Jackendoff (1977), em que ele descreve a estrutura do inglês de acordo com o modelo X-barra, o número máximo de barras é 3. No trabalho de Lemle (1984), a autora assumia como número máximo de barras ora 2, ora 3, ora 4, sem estabelecer um *n* máximo limite. Outros estudos importantes que se utilizaram da teoria X-barra, como Williams (1975), Emmonds (1976) e Stuurman (1985), postularam um número máximo diferente, ou sequer respeitaram o princípio da uniformidade, apresentando projeções máximas de diferentes núcelos com diferentes números máximos de barra.

Maximalidade é o princípio que postula que todos os elementos funcionais também devem ser projeções máximas. Esse princípio é talvez o mais controverso, já que

pode "complicar" a análise e a descrição sintática de estruturas sintagmáticas de superfície[35]. Alguns estudos para o português, como Lemle (1984), Lobato (1986), Souza e Silva & Koch (1993) e Othero (2006), não seguiram esse princípio e utilizaram rótulos como "determinante", "pré-determinante", etc., que não obedeciam ao modelo X-barra. Ou seja, esses elementos não apresentavam a estrutura interna XP – X' – X^0. De acordo com **maximalidade**, devemos entender o esquema X-barra que apresentamos acima como o seguinte:

Onde XP, YP, WP e ZP são projeções.

Por fim, **binaridade** é um princípio que regula a organização esquemática da X-barra. Repare que as ramificações estruturais na árvore sintática são sempre binárias, o que irá acarretar que nossas regras de reescrita também deverão ser sempre binárias. Para Santorini & Kroch (2007: 70), "hoje é largamente aceita a suposição de que a estrutura sintática é, no máximo, ramificada binariamente (em outras palavras a ramificação binária parece ser uma forma universal)[36]". No entanto, há casos em que a binaridade não é clara, como em estruturas coordenadas (**[João e Maria] foram ao cinema**) e na representação de verbos com mais de um argumento semântico (**João [deu-me um presente]**). Voltaremos ao assunto adiante[37].

[35] Ver, por exemplo, as críticas de Culicover & Jackendoff (2005).
[36] Trecho original: "it is now widely assumed that syntactic structure is at most binary-branching (in other words, binary-branchingness is assumed to be a formal universal)".
[37] Ainda sobre a binaridade, cf. Kayne (1984) e Kornai & Pullum (1990).

2.3 AS PROJEÇÕES MÁXIMAS EM NOSSAS ANÁLISES

Iremos agora mostrar quais serão as projeções máximas que adotamos em nossas análises da estrutura sintagmática do português. Comecemos pelos cinco núcleos lexicais: NP, VP, AP, AdvP e PP[38].

Há boas evidências na literatura de que existam pelo menos estas quatro projeções máximas cujo núcleo seja um elemento lexical: NP, VP, AP e PP. Além da vasta literatura em sintaxe gerativa que adota essas quatro projeções máximas, outros modelos e formalismos em sintaxe também costumam adotar esses quatro rótulos em suas análises, desde estudos gramaticais mais "tradicionais" (Luft, 1986; e Azeredo, 2000, por exemplo) até modelos mais recentes como o *Simpler Syntax*, a HPSG e a Teoria da Otimidade.

Os núcleos lexicais são identificados pela combinação de categorias lexicais de dois traços distintivos fundamentais, nominal [N] e verbal [V], que podem ter dois valores, positivo (+) ou negativo (-). Combinando os valores, obtemos a seguinte tabela[39]:

	[+N]	**[-N]**
[-V]	Nome	Preposição
[+V]	Adjetivo	Verbo

Tabela 2.1: Núcleos lexicais

Chamamos essas projeções de lexicais porque seus núcleos podem **subcategorizar** seus complementos. Essas quatro classes gramaticais têm seu *status* relativamente garantido entre os estudos gramaticais. No entanto, repare que não há lugar para o advérbio na tabela acima, ainda que, acreditamos, o advérbio também possa subcategorizar seus complementos. Mateus & Xavier (1992: 611), por exemplo, dizem que o sintagma adverbial é a

> projecção máxima de um advérbio. É, geralmente, constituído apenas pelo advérbio ou por este e um especificador também adverbial, funcionando como quantificador. O advérbio pode seleccionar um complemento. Exemplo: independentemente de tudo.

[38] Mioto et al. (2004) defendem a ideia de que existem **preposições lexicais** e **preposições funcionais**. Essa distinção não será relevante para nosso estudo.
[39] Essa classificação data de Chomsky (1970).

Lemle (1984: 130) também concorda com a postulação da projeção máxima adverbial: "um advérbio pode ter complementos, e um advérbio mais os seus complementos perfaz um sintagma adverbial". Diversas outras propostas de descrição gramatical igualmente propõem a existência do sintagma adverbial, como Jackendoff (1977), Huddleston (1984), Quirk et al. (1985), Aarts & Haegeman (2006), etc[40].

Adotamos o advérbio como núcleo de uma projeção máxima em trabalhos anteriores (Othero, 2006; Menuzzi & Othero, 2008) e continuaremos a fazê-lo aqui. Em resumo, adotaremos estes cinco núcleos lexicais, **N**, **P**, **A**, **V** e **Adv**, cada um deles com sua projeção máxima, **NP**, **PP**, **AP**, **VP** e **AdvP**.

Antes de prosseguir, seremos obrigados a fazer um *mea culpa*: não iremos nos aprofundar na definição exata de cada um desses núcleos lexicais. Acreditamos que um estudo que pretenda definir cada uma dessas classes gramaticais por si só poderia trazer trabalho suficiente para render um livro[41]. Por isso, seguiremos o consenso entre os estudos gramaticais do português sempre que possível e adotaremos as definições que encontramos ali. Quando lidarmos com casos problemáticos (como nas seções 4.3 e 4.4, em que trataremos do sintagma preposicional e do sintagma adverbial, respectivamente), faremos algumas escolhas, tomaremos algumas posições e argumentaremos em favor delas. O mesmo irá valer sobre os núcleos funcionais de que falaremos a seguir.

Os núcleos funcionais têm função essencialmente sintática e não subcategorizam seus complementos[42]. Na verdade, em nosso trabalho, a distinção entre núcleos lexicais e núcleos funcionais não será relevante. O importante é estipularmos quais tipos de núcleos e de projeções máximas iremos adotar na descrição dos sintagmas e da sentença em português e justificar sua importância com a descrição sintática que iremos apresentar.

Uma projeção máxima que utilizaremos em nossas análises será a projeção máxima do **DP**, o sintagma determinante (do inglês *determiner phrase*), projeção máxima do núcleo **determinante**. Desde o final da década de 1980, graças principalmente ao trabalho de Abney (1987), o antigo NP passou a ser entendido como uma projeção do

[40] Winograd (1983: 53) também "aceita" o advérbio como categoria gramatical legítima na língua (inglesa), ainda que com uma advertência: "[o advérbio] é a sobra entre as categorias (*leftover category*) – qualquer coisa que não se encaixa em outra classe recebe o nome de advérbio"!

[41] O assunto da categorização das classes de palavras é tratado por diversos autores – para uma introdução ao problema, ver Assunção (2001), Azeredo (2001), Perini (2003, 2006) e Aarts & Haegeman (2006); para um estudo linguístico pioneiro em português do Brasil, ver Câmara Jr. (1970); para uma tentativa bastante ampla e interessante, ver Perini (1989, 2000).

[42] Uma diferença interessante entre núcleos lexicais e funcionais tem caráter semântico: de acordo com Glanzberg (2008), os núcleos funcionais não podem ser interpretados metaforicamente, enquanto os núcleos lexicais podem.

determinante. Em nossa descrição do sintagma nominal, adotaremos a projeção máxima DP, aceitando, então, a **hipótese do DP**[43]. Assumiremos que a posição do D em português pode ser ocupada por **pronomes demonstrativos**, **pronomes pessoais**, **artigos**, **quantificadores**, **elementos-QU** ou por um **elemento vazio** (entraremos em mais detalhes na seção 4.1).

Ainda na descrição do sintagma nominal, adotaremos os núcleos **Q** (quantificador), **Poss** (possessivo) e **Num** (numeral), que projetam **QP**, **PossP** e **NumP**, respectivamente (cf. 4.1). Nossas propostas de análises dos agrupamentos nominais do português se baseiam boa parte em Abney (1987), Radford (1997a,b) e Zamparelli (2000), que apresentam análises para o inglês; e Lemle (1984), Luft (1986), Perini (1989, 2000), Raposo (1992) e Mioto et al. (2004), com suas análises para o português. Seguimos de perto também a proposta de David (2007) e Othero & David (no prelo), que descrevem a estrutura do DP visando à implementação computacional.

Para o estudo da estrutura e descrição da **sentença**, ao contrário do que fizemos em trabalhos anteriores, não iremos utilizar o símbolo **S**. Como vimos anteriormente, uma regra tal como (2.9) violaria alguns princípios da teoria X-barra, como o princípio da endocentricidade:

(2.9) S → NP VP[44]

Para a descrição da sentença em português, adotaremos as duas projeções máximas **IP** (sintagma flexional, do inglês *inflectional phrase*); e **CP** (sintagma complementizador, do inglês *complementizer phrase*), seguindo modelos correntes de descrição gramatical que também utilizam o formalismo da teoria X-barra. Um breve histórico da evolução do estudo da sentença, de **S** até as projeções **IP** e **CP**, pode ser vista em Falk (2006), de quem tomamos emprestados alguns trechos:

> A sentença como uma categoria sui generis. Em sintaxe gerativa, a categoria S é analisada expandindo-se como uma seqüência de NP-VP

[43] Cf. Abney (1987) e Coene & D'Hulst (2003a, 2003b), entre outros. As análises com o DP são mais ou menos consensuais em pesquisa sintática atual, dentro da Teoria dos Princípios e Parâmetros (cf., no entanto, Cinque, 1995, e Escribano, 2006, para críticas à hipótese do DP).

[44] Tal regra é comumente encontrada em manuais de sintaxe antigos para a descrição da sentença em português. Nós mesmos implementamos computacionalmente essa regra na gramática do *parser* que apresentaremos no próximo capítulo, o Grammar Play. Na verdade, encarar a sentença como o símbolo inicial **S** era uma prática comum até Chomsky (1986b), que substitui S' por CP (*Complementizer Phrase*) e S por IP (*Inflectional Phrase*), cada um com sua estrutura interna respeitando lexicalidade e endocentricidade (IP – I' – I(nfl); CP – C' – C(omp)), como veremos em seguida.

(...).
Chomsky (1957) caracteriza o componente de regras sintagmáticas como sendo "definido por um conjunto finito Σ de seqüências iniciais e um conjunto finito F de 'fórmulas de instrução' com a forma X → Y..." [Chomsky, 1957: 29], e descreve as gramáticas sintagmáticas como gramáticas [Σ, F]. O conceito do símbolo inicial é então considerado parte da definição do componente sintagmático da gramática. O símbolo inicial é a sentença, como se percebe pela discussão sobre se o símbolo Σ deve consistir apenas pela Sentença ou se devem ser incluídas categorias distintas como Sentença Declarativa, Sentença Interrogativa, etc. Uma outra indicação do status especial da sentença/oração nesse trabalho é o tratamento da recursão. Enquanto a recursão era geralmente codificada nas regras de estrutura sintagmática, a recursão oracional era analisada pela geração de sentenças separadas e então encaixadas uma na outra por uma transformação generalizada.
Em Chomsky (1965), o uso de um mecanismo formal distinto para a recursão frasal foi abandonado, mas a singularidade da sentença foi mantida.
Chomsky (1970), que introduziu a teoria X', declarou que a 'regra inicial' da base era S → N'' V'', sem nenhuma proposta de incorporar a sentença ao sistema X-barra. (...)
Jackendoff (1977), em sua tentativa de generalizar estruturalmente o sujeito-da-sentença com o possuidor-do-NP (assumindo que o possuidor fosse especificador do NP), propõe que uma sentença seja a projeção máxima do V. VP é uma categoria submáxima, isto é, não é uma oração plena. (Os auxiliares também são especificadores).
Gazdar, Pullum & Sag (1982) apresentam uma análise semelhante à de Jackendoff (sentença = V'', "VP" = V'). A única diferença é que eles analisam auxiliares como sendo verbos. (...)
A sentença como projeção máxima da Inflexão (I). Chomsky (1981) menciona essa possibilidade. (...)
Chomsky (1986b) popularizou essa análise (...). Essa implementação tem a vantagem de fornecer uma estrutura X' plena.45

[45] Trecho original: "Sentence is a sui generis category. In generative syntax this has been realized as analyzing the special category S as expanding as a subject-predicate sequence of NP-VP (…).
Chomsky (1957) characterizes the phrase structure rule component of the grammar as being 'defined by a finite set Σ of initial strings and a finite set F of 'instruction formulas' of the form X → Y…', and describes phrase structure grammars as [Σ, F] grammars. The concept of an initial symbol is thus taken to be part of the definition of the phrase structure component of the grammar. This initial symbol is the sentence, as is made clear by the discussion over whether the set Σ consists just of Sentence or might include distinct categories Declarative Sentence, Interrogative Sentence, etc. Another indication of the special status of the sentence/clause in this early work is the treatment of recursion. While recursion was generally encoded in the phrase structure rules, clausal recursion was analyzed by generating separate sentences, and then embedding one in the other by a generalized transformation.
In Chomsky (1965) the use of a distinct formal device for clausal recursion was abandoned, but the uniqueness of the sentence was retained.
Chomsky (1970), which introduced X' theory, stated the 'initial rule' of the base to be S → N'' V'', with no attempt to incorporate the sentence into the X' system. (…)
Jackendoff (1977), trying to structurally generalize subject-of-sentence with possessor-of-NP (assuming the possessor is specifier of NP), proposes that sentence is the maximal projection of V. VP is a submaximal category, i.e. not a full phrase. (Auxiliaries are also specifiers.)
Gazdar, Pullum, & Sag (1982) have an analysis similar to Jackendoff's (sentence=V'', 'VP'=V'). The only diTerence is that they analyze auxiliaries as verbs. (…)
Sentence is I^{max}. Chomsky (1981) mentions this as a possibility. (…)

A fim de propormos regras gramaticais do português afinadas com o modelo da teoria X-barra, iremos adotar as projeções IP e CP. Tentaremos formalizar propostas interessantes de descrição sintática da estrutura da frase em português (como Pontes, 1973; Lobato, 1975; Lemle, 1984; Luft, 1986; Perini, 1989, 2000; e Othero, 2006), **ampliando** seu poder de descrição em alguns casos, **atualizando** as análises em outros casos, utilizando, para esse fim, o modelo proposto pela teoria X-barra.

Chomsky (1986) popularized this analysis (...). This implementation has the advantage of providing a full X' structure".

3 O GRAMMAR PLAY E AS ANÁLISES DE OTHERO (2004)

Only God can make a tree[46]

Joyce Kilmer

Em Othero (2004), apresentamos uma análise da sentença simples do português e uma aplicação computacional dessa análise. Desenvolvemos um programa de *parsing* sintático, o *parser* Grammar Play. O Grammar Play analisa unicamente sentenças simples declarativas do PB, ou seja, sentenças que contenham apenas um verbo e que não sejam interrogativas[47]. Ele é capaz de dizer quais dessas sentenças são gramaticais e quais não são, além de atribuir às sentenças gramaticais uma estrutura sintagmática baseando-se no modelo da teoria X-barra tradicional. Além disso, a interface gráfica do Grammar Play permite que o usuário visualize a estrutura das sentenças tanto em formato de colchetes rotulados quanto em forma de um marcador sintagmático (ou árvore sintática). Abaixo temos o exemplo de uma sentença analisada pelo Grammar Play[48]:

[46] "Somente Deus pode fazer uma árvore".

[47] Na verdade, ficaram de fora apenas as interrogativas-QU, do tipo *Quem é amigo da Maria? O que a Maria quer de presente?*, etc. Sentenças interrogativas do tipo sim-ou-não, que não alteram sua estrutura com relação à sua correspondente declarativa, não apresentam problemas. Compare as estruturas dos pares *O João é amigo da Maria. / O João é amigo da Maria?*; *A Maria quer chocolates de presente. / A Maria quer chocolates de presente?*.

[48] O Grammar Play está disponível para *download* em http://www.geocities.com/ gabriel_othero/.

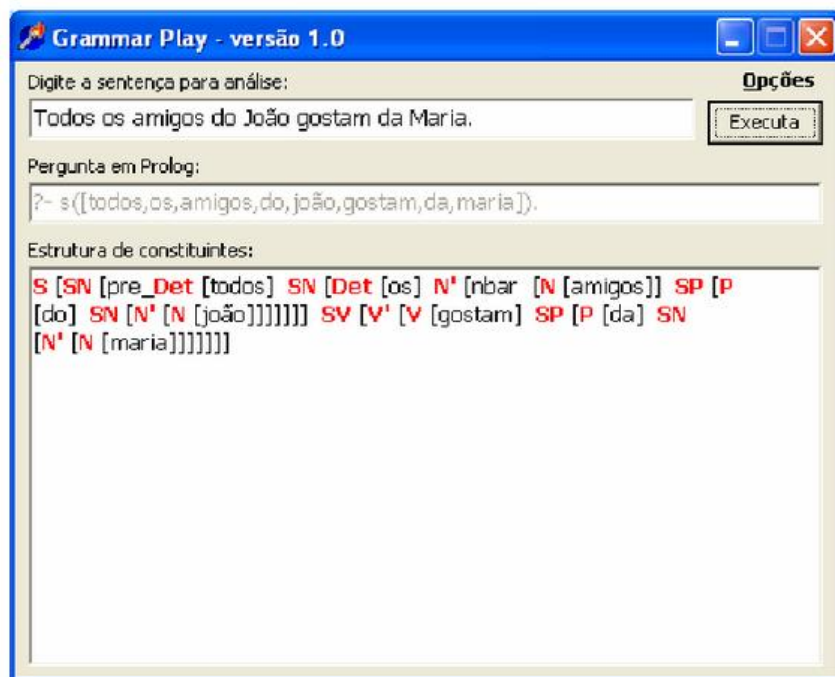



Figura 3.1: Exemplo de sentença analisada com colchetes rotulados

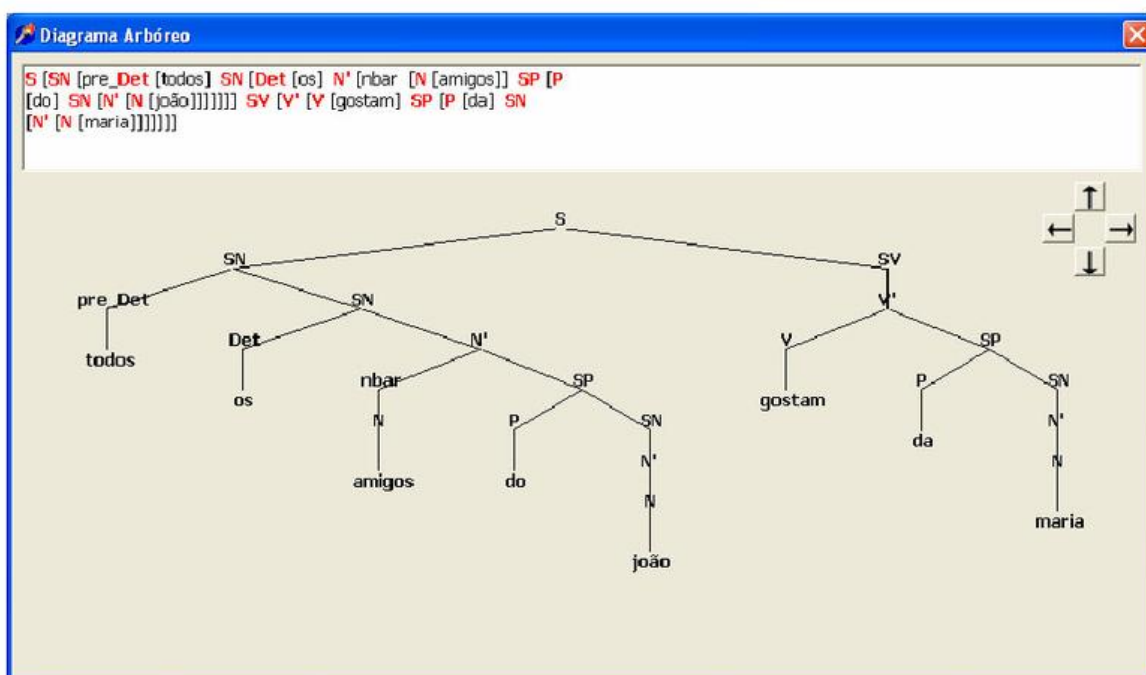



Figura 3.2: Exemplo de sentença analisada com sua estrutura arbórea

As duas ideias centrais subjacentes ao desenvolvimento do Grammar Play foram (i) a de apresentar um programa de *parsing* sintático para o PB que fosse *user-friendly* (desenvolvemos o programa pensando em todo aquele nosso colega linguista que não tem noções de programação, mas que deseja saber mais sobre a interface da Sintaxe com a

computação); e (ii) a de implementar computacionalmente uma descrição do português de acordo com uma teoria linguística consistente (em nosso caso, uma versão da teoria X-barra).

Para pegar um trecho emprestado de Menuzzi & Othero (2008: 1):

> O parser [Grammar Play] é uma primeira tentativa de expandir a cobertura de analisadores semelhantes, como o esboçado em Pagani (2004) e Othero (2004). Os objetivos que guiam a presente versão do Grammar Play são o de implementar computacionalmente modelos lingüísticos coerentes aplicados à descrição do português e o de criar uma ferramenta computacional que possa ser usada didaticamente em aulas de introdução à sintaxe e lingüística, por exemplo.

As regras da gramática do *parser* foram elaboradas a partir de regras relativamente simples já consagradas na literatura pertinente sobre descrição sintática do português e a partir da análise de exemplos que foram encontrados em livros de introdução à teoria sintática. A gramática do Grammar Play desde o início seguiu o modelo de uma gramática PSG livre de contexto enriquecida com traços lexicais. Tentamos seguir os moldes do esquema apresentado pela teoria X-barra tradicional. Todas as regras foram adaptadas à linguagem Prolog, com o recurso das *Definite Clause Grammars* (DCGs)[49].

As regras de descrição sintática que apresentamos foram as seguintes[50]:

(1) S → NP VP

(2) S → VP NP

(3) S → VP

(4) S → AdvP S

(5) NP → det N'

(6) NP → N'

(7) NP → pré-det NP

(8) N' → Pro[51]

(9) N' → N

[49] Julgamos não ser pertinente discutir o recurso das DCGs aqui. Sobre o assunto, remetemos o leitor a Pereira & Warren (1980), Pereira & Shieber (1987), Bratko (1997) e Brna (2001).

[50] Estamos apresentando essas regras porque partiremos delas para efetuarmos nossas análises no capítulo 4.

[51] "Pro" é um pronome. Não confundir com **pro** ("prozinho") nem com **PRO** ("prozão").

(10) N’ → AP N’

(11) N’ → N’ AP

(12) N’ → N’ PP

(13) AP → Adj’ AdvP

(14) AP → Adj’ PP

(15) AP → AdvP Adj’

(16) AP → Adj’

(17) Adj’ → Adj

(18) Adj’ → AdvP Adj’

(19) Adj’ → Adj’ AdvP

(20) Adj’ → Adj’ PP

(21) PP → P NP

(22) PP → P AdvP

(23) VP → V’

(24) VP → V’ PP

(25) VP → AdvP V’

(26) VP → V’ AdvP

(27) V’ → V’ NP

(28) V’ → V’ PP

(29) V’ → AdvP V’

(30) V’ → V’ AdvP

(31) V’ → V

(32) V’ → V NP

(33) V’ → V PP

(34) V’ → V AP

(35) V’→ V AdvP

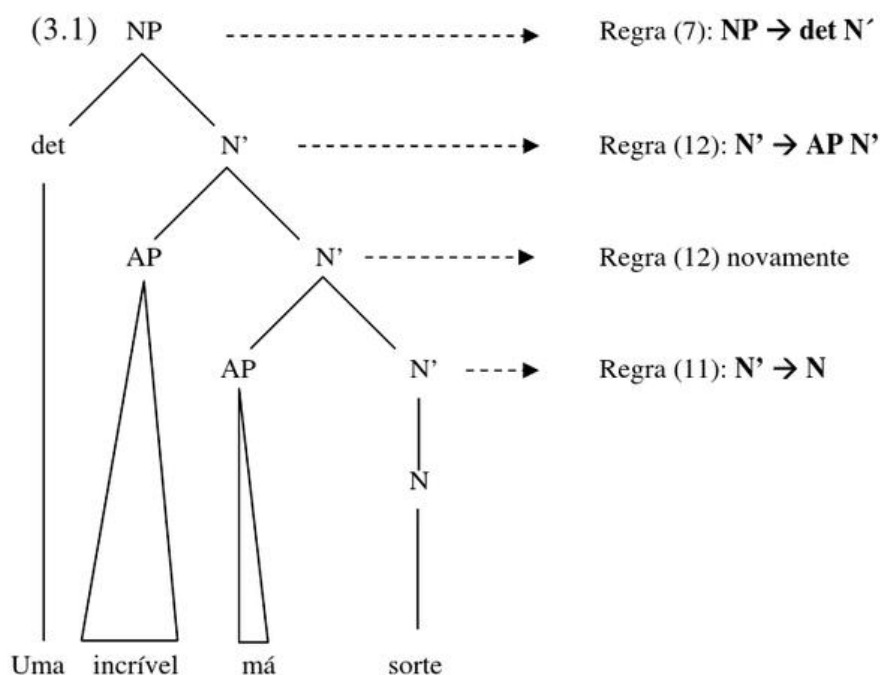
(36) AdvP → Adv'

(37) AdvP → AdvP Adv'

(38) AdvP → Adv' PP

(39) Adv' → Adv

(40) Adv' → Adv' PP

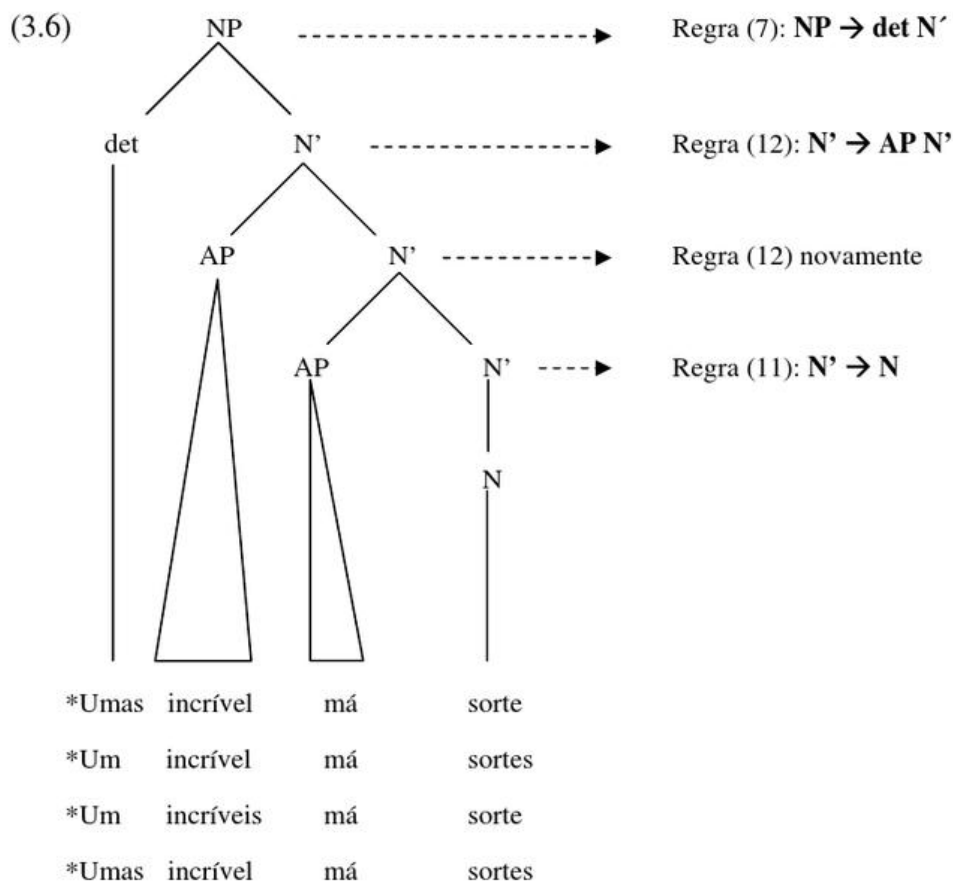
Essas 40 regras de descrição da sentença simples em PB resultaram em 67 regras implementadas na gramática do Grammar Play. Essa diferença no número total de regras se deve principalmente ao fator **hipergeração**. Ou seja, as regras de descrição sintática que elaboramos estão teoricamente corretas, mas elas hipergeram estruturas e hipergeneralizam as análises do *input* linguístico que o *parser* recebe. As regras (1) a (40) são o que chamamos de **macrorregras** sintáticas. E essa noção será importante para o desenvolvimento de nosso trabalho no próximo capítulo. Por exemplo, poderíamos analisar o NP **[uma incrível má sorte]**, utilizando as regras (7), (12) e (11) – sem entrar em detalhes na estrutura do AP. Confira:

Porém, essas mesmas regras poderiam gerar\analisar os NPs agramaticais (3.2), (3.3), (3.4) e (3.5), como mostramos em (3.6):

(3.2) *Umas incrível má sorte

(3.3) *Um incrível má sortes

(3.4) *Um incríveis má sorte

(3.5) *Umas incrível má sortes

(3.6)

O problema é que essas macrorregras sintáticas, tais como apresentamos de (1) a (40), não preveem a **unificação** (ou checagem) **de traços lexicais**, hipergerando estruturas que são agramaticais na língua. Isso foi algo com que tivemos de nos preocupar em nossa implementação computacional da gramática em Prolog. Daí as 40 regras de descrição sintática da sentença simples do português se transformarem em 67 regras implementadas computacionalmente em Prolog. Todo esse trabalho nos lembrou uma frase de Ray Jackendoff (2002: 167): "o problema é que a sintaxe bonitinha sozinha não tem tanto uso para a compreensão da máquina"[52].

Na verdade, o problema já é bem conhecido. David (2007: 13), em seu trabalho de

[52] Trecho original: "The problem is that fancy syntax alone isn't that much use for machine understanding".

implementar computacionalmente (também em Prolog) um fragmento de gramática do DP, por exemplo, reparou que

> alguns autores, como Raposo (1992), entendem que, para gerar um fragmento de gramática de estrutura sintagmática do português, faz-se necessário que se determinem regras capazes de gerar e reconhecer essas expressões nominais como bem formadas ou gramaticais. Assim é que Raposo (1992) apresenta regras de reescrita categoriais, as quais não conseguem assegurar que os traços gramaticais, que diferentes palavras trazem em sua estrutura, possam ser compatíveis com aqueles de outras palavras da mesma expressão.

Por isso, tivemos de enriquecer os itens lexicais do *parser* com traços gramaticais que percolavam, de certa forma, os nós dos constituintes, desde os itens terminais até o nó máximo da sentença. Assim, garantimos a análise de boa formação dos sintagmas e da sentença.

Por exemplo, os verbos contêm a informação de **número** e de **subcategorização sintática**; e os substantivos, adjetivos e determinantes, a informação de **número** e **gênero**[53]. Essas informações presentes nos itens lexicais percolam os nós até o nó máximo, efetuando uma **unificação de traços**. Abaixo, mostramos a entrada de alguns itens lexicais no léxico do Grammar Play, para dar uma ideia clara do que estamos falando:

[53] Evidentemente, o léxico pode conter outros tipos de informação, além desses traços. Outros dois trabalhos de implementação sintática em Prolog de fragmentos do PB, como Pagani (2004) e David (2007), trazem diferentes traços lexicais, como gênero, número, pessoa, definitude, etc.

```
% Verbos Intransitivos (i)
v(sing, i, morre).
v(plur, i, morrem).

% Verbos Transitivos Diretos (td)
v(sing, td, come).
v(plur, td, comem).

% Verbos Transitivos Indiretos (ti)
v(sing, ti(de), precisa).
v(plur, ti(de), precisam).

% Verbos Bitransitivos (tdi)
v(sing, tdi(em), coloca).
v(plur, tdi(em), colocam).

% Verbos de Ligação (vl)
v(sing, vl, continua).
v(plur, vl, continuam).
```

Quadro 3.1: Verbos no léxico do Grammar Play

```
% Substantivos
n([fem,sing], amiga).
n([fem,plur], amigas).
n([masc,sing], amigo).
n([masc,plur], amigos).

% Adjetivos
adj([fem,sing], alta).
adj([fem,plur], altas).
adj([masc,sing], alto).
adj([masc,plur], altos).
```

Quadro 3.2: Substantivos e adjetivos no léxico do Grammar Play

Esses traços que aparecem entre parênteses nos itens lexicais estão presentes também nas regras de reescrita que implementamos na gramática do *parser*. Com uma descrição sintática "limpa" ou "pura", nos moldes da teoria X-barra, como costumamos encontrar em manuais de descrição sintática do português[54], acabamos formulando macrorregras gramaticais.

Abaixo, apresentamos as regras sintagmáticas que desenvolvemos em nossa gramática em Prolog. Repare como as informações dos traços morfossintáticos presentes

[54] Como, por exemplo, Lemle (1984), Lobato (1986), Raposo (1992), Mioto et al. (2004), entre outros.

no léxico também aparecem nas regras categoriais de reescrita.

	% Sentença (S)
1)	s([sn, NP, sv, VP]) --> sn([_,Num], NP), sv(Num, VP).
2)	s([sv, VP, sn, NP]) --> sv(Num, VP), sn([_,Num], NP).
3)	s([sv, SVimp]) --> svimp(_, SVimp).
4)	S([sadv, AdvP, s, S]) --> sadv(AdvP), s(S).

Quadro 3.3: Regras da sentença

5)	sn(Conc, [det, [Det], n_bar, N_Barra]) --> det(Conc, Det), n_barra(Conc, N_Barra).
6)	sn(Conc, [n_bar, N_Barra]) --> n_barra(Conc, N_Barra).
7)	sn(Conc, [pre_det, [Pre_Det], sn, NP]) --> pre_det(Conc, Pre_Det), sn(Conc, NP).
8)	n_barra(Conc, [pro, [Pro]]) --> pro(Conc, Pro).
9)	n_barra(Conc, [n, [N]]) --> n(Conc, N).
10)	n_barra(Conc, [sadj, AP, n_bar, N_Barra]) --> sadj(Conc, AP), n_barra(Conc, N_Barra).
11)	n_barra(Conc, [n_bar, X, sadj, AP]) --> x(Conc, X), sadj(Conc, AP).
12)	x(Conc, [n, [N]]) --> n(Conc, N).
13)	x(Conc, [n_bar, Y, sadj, AP]) --> y(Conc, Y), sadj(Conc, AP).
14)	y(Conc, [n, [N]]) --> n(Conc, N).
15)	y(Conc, [n_bar, Z, sadj, AP]) --> z(Conc, Z), sadj(Conc, AP).
16)	z(Conc, [n, [N]]) --> n(Conc, N).
17)	n_barra(Conc, [nbar, X, sp, PP]) --> x(Conc, X), sp(P, PP).

Quadro 3.4: Regras do NP

18)	sadj(Conc, [adj_bar, Adj_Barra, sadv, AdvP]) --> adj_barra(Conc, Adj_Barra), sadv(AdvP).
19)	sadj(Conc, [adj_bar, Adj_Barra, sp, PP]) --> adj_barra(Conc, Adj_Barra), sp(P, PP).
20)	sadj(Conc, [sadv, AdvP, adj_bar, Adj_Barra]) --> sadv(AdvP), adj_barra(Conc, Adj_Barra).
21)	sadj(Conc, [adj_bar, Adj_Barra]) --> adj_barra(Conc, Adj_Barra).
22)	adj_barra(Conc, [adj, [Adj]]) --> adj(Conc, Adj).
23)	adj_barra(Conc, [sadv, AdvP, adj_bar, Adj_Barra]) --> sadv(AdvP), adj_barra(Conc, Adj_Barra).
24)	adj_barra(Conc, [adj_bar, A, sadv, AdvP]) --> a(Conc, A), sadv(AdvP).
25)	adj_barra(Conc, [adj_bar, A, sp, PP]) --> a(Conc, A), sp(P, PP).
26)	A(Conc, [adj, [Adj]]) --> adj(Conc, Adj).

Quadro 3.5: Regras do AP

27)	sp(Lex, [p, [P], sn, NP]) --> p(Lex, inf, P), sn(Conc, NP).
28)	sp(Lex, [p, [P], sn, NP]) --> p(Lex, Conc, P), sn(Conc, NP).

Quadro 3.6: Regras do PP

30)	sv(Num, [v_bar, V_Barra, sadv, AdvP]) --> v_barra(Num, i, V_Barra), sadv(AdvP).
31)	sv(Num, [v_bar, V_Barra, sp, PP]) --> v_barra(Num, i, V_Barra), sp(_, PP).
32)	sv(Num, [v_bar, V_Barra]) --> v_barra(Num, _, V_Barra).
33)	svimp(Num, [v_bar, V_Barra_i]) --> v_barra_i(Num, _, V_Barra_i).
34)	sv(Num, [sadv, AdvP, v_bar, V_Barra]) --> sadv(AdvP), v_barra(Num, _, V_Barra).
35)	svimp(Num, [sadv, AdvP, v_bar, V_Barra_i]) --> sadv(AdvP), v_barra_i(Num, _, V_Barra_i).
36)	sv(Num, [v_bar, V_Barra, sadv, AdvP]) --> v_barra(Num, _, V_Barra), sadv(AdvP).
37)	svimp(Num, [v_bar, V_Barra_i, sadv, AdvP]) --> v_barra_i(Num, _, V_Barra_i), sadv(AdvP).
38)	sv(Num, [v_bar, V_Barra, sn, NP]) --> v_barra(Num, _, V_Barra), sn(_, NP).
39)	sv(Num, [v_bar, V_Barra, sp, PP]) --> v_barra(Num, _, V_Barra), sp(_, PP).
40)	svimp(Num, [v_bar, V_Barra_i, sp, PP]) --> v_barra_i(Num, i, V_Barra_i), sp(_, PP).
41)	v_barra(Num, _, [sadv, AdvP, v_bar, V_Barra]) --> sadv(AdvP), v_barra(Num, _, V_Barra).
42)	v_barra_i(Num, _, [sadv, AdvP, v_bar, V_Barra_i]) --> sadv(AdvP), v_barra_i(Num, _, V_Barra_i).
43)	v_barra(Num, i, [v, [V]]) --> v(Num, i, V).
44)	v_barra(Num, td, [v, [V], sn, NP]) --> v(Num, td, V), sn(_, NP).
45)	v_barra(Num, ti, [v, [V], sp, PP]) --> v(Num, ti(P), V), sp(P, PP).
46)	v_barra(Num, vl, [vl, [V], sn, NP]) --> v(Num, vl, V), sn(Conc, NP).
47)	v_barra(Num, vl, [vl, [V], sadj, AP]) --> v(Num, vl, V), sadj(Conc, AP).
48)	v_barra(Num, vl, [v, [V], sp, PP]) --> v(Num, vl, V), sp(_, PP).
49)	v_barra(Num, vl, [v, [V], sadv, AdvP]) --> v(Num, vl, V), sadv(AdvP).
50)	v_barra_i(_, i, [v, [Vimp]]) --> vimp(_, i, Vimp).
51)	v_barra_i(_, td, [v, [Vimp], sn, NP]) --> vimp(_, td, Vimp), sn(_, NP).
52)	v_barra_i(_, _, [v_bar, T, sadv, AdvP]) --> t(_, _, T), sadv(AdvP).
53)	v_barra(Num, _, [v_bar, T, sadv, AdvP]) --> t(Num, _, T), sadv(AdvP).
54)	v_barra(Num, _, [v_bar, T, sp, PP]) --> t(Num, _, T), sp(_, PP).
55)	v_barra(Num, _, [v_bar, T, sn, NP]) --> t(Num, _, T), sn(_, NP).
56)	t(Num, _, [v, [V]]) --> v(Num, _, V).
57)	t(Num, _, [v, [V], sp, PP]) --> v(Num, _, V), sp(_, PP).
58)	t(Num, _, [v, [V], sn, NP]) --> v(Num, _, V), sn(_, NP).
59)	t(Num, _, [v_bar, TT, sadv, AdvP]) --> tt(Num, _, TT), sadv(AdvP).
60)	t(_, _, [v, [Vimp], sn, NP]) --> vimp(_, td, Vimp), sn(_, NP).
61)	tt(Num, _, [v, [V]]) --> v(Num, _, V).

Quadro 3.7: Regras do VP

```
61)    sadv([adv_bar, Adv_Barra]) --> adv_barra(Adv_Barra).
62)    sadv([sadv, B, adv_bar, Adv_Barra]) --> b(B), adv_barra(Adv_Barra).
63)    sadv([adv_bar, Adv_Barra, sp, PP]) --> adv_barra(Adv_Barra), sp(P, PP).
64)    adv_barra([adv, [Adv]]) --> adv(Adv).
65)    adv_barra([adv_bar, C, sp, PP]) --> c(C), sp(P, PP).
66)    b([adv_bar, Adv_Barra]) --> adv_barra(Adv_Barra).
67)    c([adv, [Adv]]) --> adv(Adv).
```

Quadro 3.8: Regras do AdvP

Não entraremos em detalhes sobre a implementação das regras sintáticas na gramática do *parser* porque não iremos propor a implementação computacional das regras que iremos apresentar aqui (cf. Othero, 2006, para detalhes da formulação e implementação computacional dessas regras gramaticais). A partir do próximo capítulo, iremos descrever os sintagmas do PB e algumas estruturas sentenciais com **a formulação de macrorregras de descrição sintática**. Em outras palavras, a gramática das estruturas sintagmáticas do PB que iremos propor não irá falar nada sobre o papel de traços lexicais. Sabemos que a descrição do léxico é fundamental para que possamos, por exemplo, implementar as regras sintáticas em ambiente computacional de maneira mais eficiente, como no exemplo de uma gramática em Prolog, que vimos acima. No entanto, iremos nos ater ao estudo das regras sintagmáticas de boa formação dos sintagmas e de algumas estruturas da sentença em PB e apenas daremos sugestões sobre os traços lexicais. As regras que apresentaremos a partir do próximo capítulo – por mais detalhadas que estejam e por mais que tenhamos pensado sobre elas – serão, ainda assim, macrorregras, que certamente estarão sujeitas à hipergeração de estruturas. No entanto, isso não é um problema, uma vez que, estando bem formadas as regras gramaticais, bastará o trabalho com os traços lexicais para que elas possam ser implementadas computacionalmente.

Ao não atrelar nossas regras à implementação computacional em uma determinada linguagem de programação (como o Prolog, por exemplo), pretendemos apresentar ao sintaticista e ao linguista computacional uma gramática da sentença do português que possa ser passível de implementação computacional em qualquer linguagem de computador apta ao trabalho em PLN. Essa gramática irá apresentar regras formais de reescrita dos sintagmas e da sentença em PB.

4 ANÁLISES DOS SINTAGMAS DO PB

When I was a graduate student, N. Chomsky would argue that writing a rule does not constitute a solution to a problem, but merely a statement of it[55].

Luigi Burzio

Dividimos este capítulo em seis grandes seções, dedicadas às quatro projeções lexicais máximas e à estrutura da sentença em PB. Primeiramente, analisamos a estrutura do sintagma nominal (NP); depois, passamos à análise do sintagma adjetival (AP), do sintagma adverbial (AdvP), do sintagma preposicional (PP) e, finalmente, do sintagma verbal (VP). Por fim, apresentamos nossas análises da estrutura da sentença (IP e CP). A sexta e última seção traz um resumo de nossas regras gramaticais.

4.1 O SINTAGMA NOMINAL

Como já mencionamos na seção 2.3, adotaremos em nossas análises o núcleo funcional **D** e sua projeção máxima, o **DP**. Para iniciar o estudo do sintagma nominal, partimos das análises que apresentamos no capítulo 3. O primeiro passo foi verificar as regras da estrutura do NP que já havíamos proposto anteriormente e tentar "atualizá-las" à nova proposta. Por conveniência, repetimos aqui as regras do sintagma nominal que havíamos apresentado:

(4.1) NP → det N'

(4.2) NP → N'

(4.3) NP → pré-det NP

(4.4) N' → Pro

(4.5) N' → N

[55] "Quando eu era um aluno de pós-graduação, Noam Chomsky dizia que escrever uma regra não era a solução de um problema; era antes a mera formulação do problema".

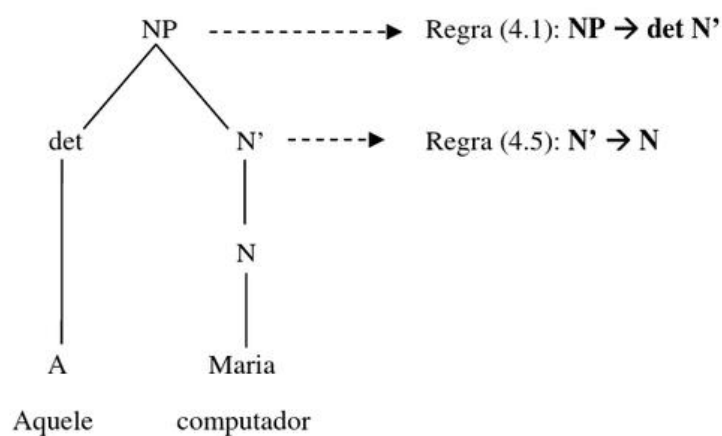
(4.6) N’ → AP N’

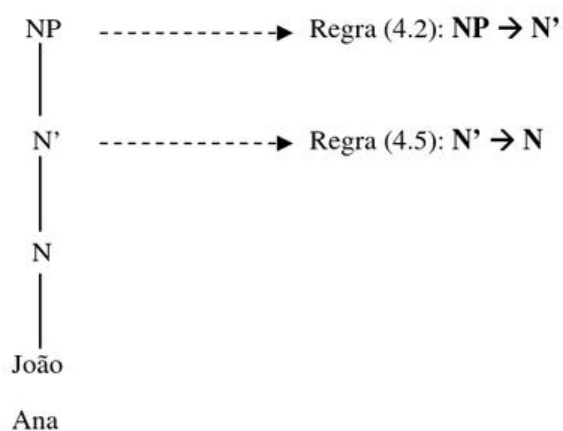
(4.7) N’ → N’ AP

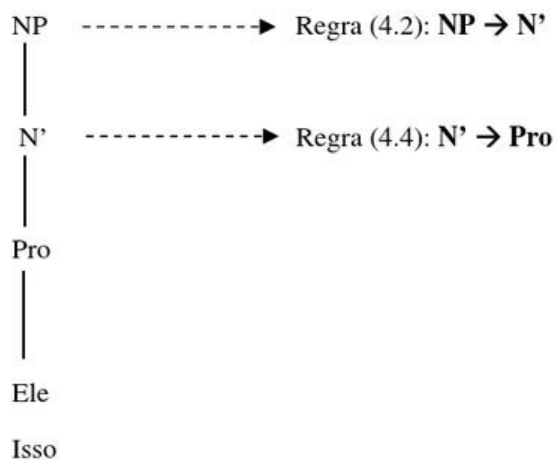
(4.8) N’ → N’ PP

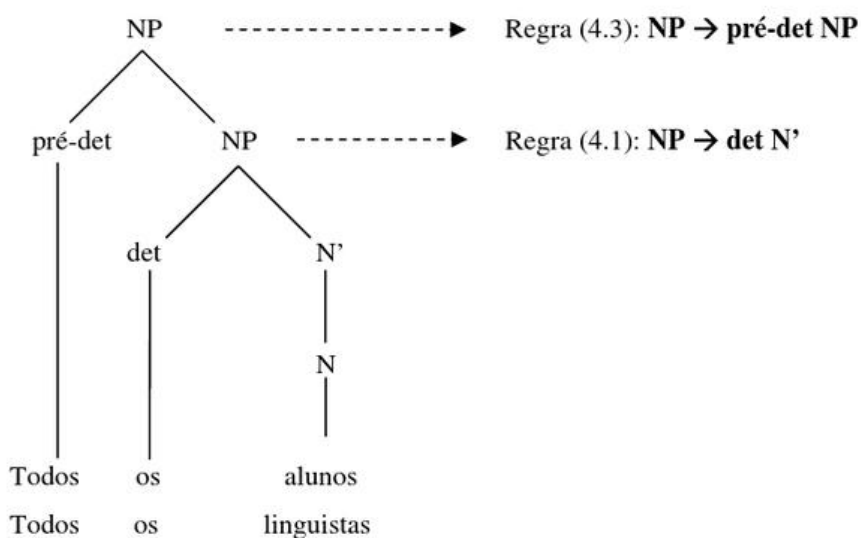
Essas oito regras gramaticais foram motivadas com base em análises de diversas sentenças e agrupamentos nominais, como os que apresentamos a seguir.

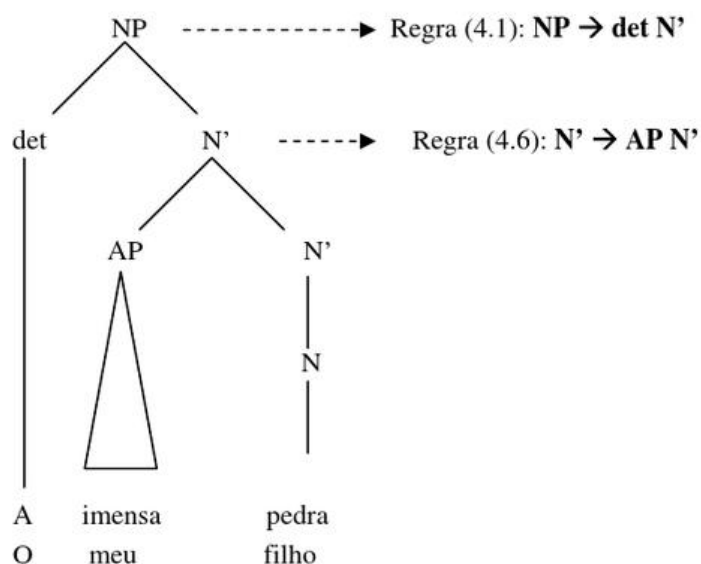
(4.9) [A Maria] adora chocolates.

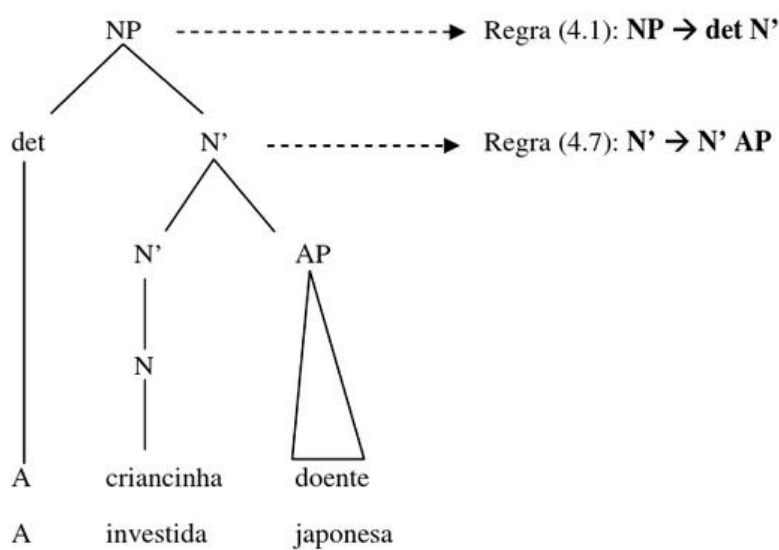
(4.10) [Aquele computador] é melhor do que este.

(4.11) [João] chegou.

(4.12) [Ana] saiu.

(4.13) [Ele] chegou.

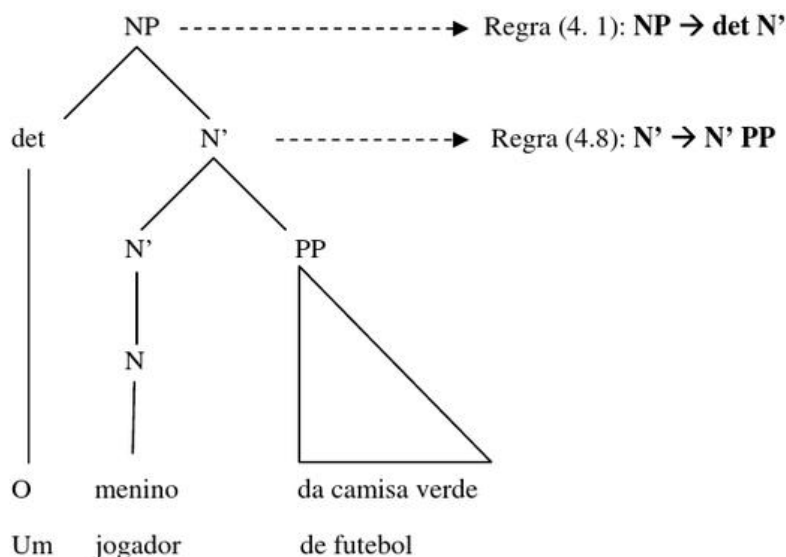
(4.14) [Isso] é extremamente importante.

(4.15) [Todos os alunos] gostaram do livro.

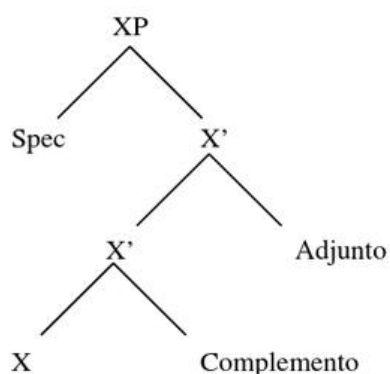
(4.16) [Todos os linguistas] já leram algum texto do Chomsky.

(4.17) [A imensa pedra] está bloqueando o caminho.

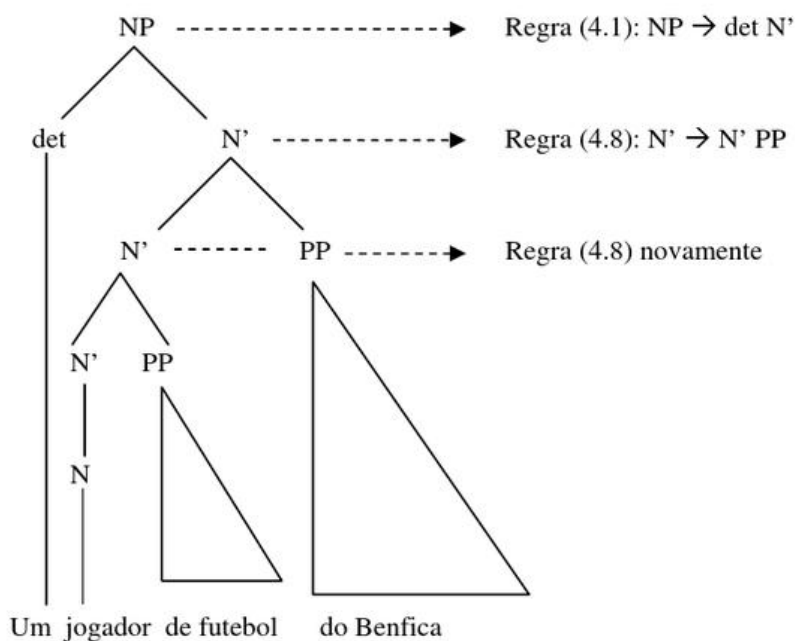
(4.18) [O meu filho] saiu.

(4.19) [A criancinha doente] melhorou.

(4.20) [A investida japonesa] foi bem sucedida.

(4.21) [O menino da camisa verde] acaba de chegar.

(4.22) [Um jogador de futebol] do Benfica56.

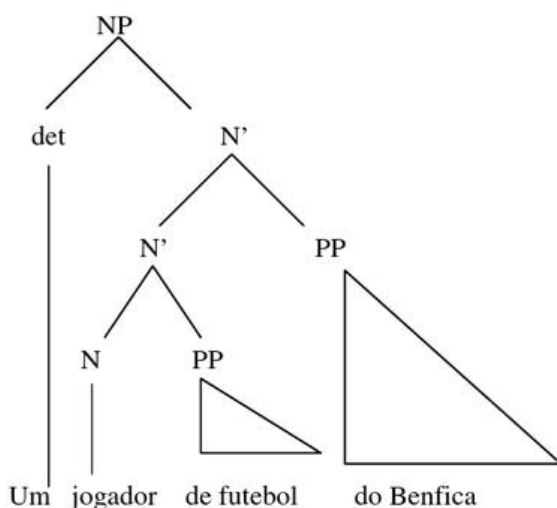
[56] Exemplo de Raposo (1992: 204). Voltaremos à discussão desse exemplo em seguida.

Essas regras nos pareceram interessantes à época e por isso resolvemos partir delas para seguir com nossas novas análises. Como vimos no capítulo 3, todas essas regras foram implementadas computacionalmente em Prolog na gramática do *parser* Grammar Play e se mostraram relativamente eficientes. Ou seja, elas estavam **formalizadas** o suficiente para que fossem implementadas em uma linguagem de programação para lidar com a sintaxe de uma língua natural. No entanto, além de não estarem afinadas com a nomenclatura corrente utilizada no modelo de descrição gramatical que adotamos, essas regras apresentam alguns problemas para a descrição dos sintagmas em português.

Por exemplo, um fator importante que o leitor pode ter percebido é que não há nenhuma regra entre (4.1) e (4.8) que preveja um complemento de N. No modelo de teoria X-barra que adotamos (cf. cap. 2), há uma diferença estrutural entre complementos e adjuntos que pode ser percebida nas relações de dominância estrutural. Reveja o esquema da X-barra abaixo:

De acordo com o modelo, o complemento está mais próximo do núcleo (é **irmão de X**), enquanto o adjunto vem ao lado do constituinte intermediário (é **irmão de X'**). Não havíamos feito essa distinção na estrutura interna do sintagma nominal nas regras que apresentamos acima. Por isso, as regras que envolvem APs e PPs mostram-nos sempre como adjuntos, ou seja, como irmãos da projeção intermediária do núcleo (**N'**), nunca como irmãos do nó terminal (**N**). Isso deve ter ficado claro especialmente na análise da sentença (4.22), de Raposo. Raposo (1992: 204) propõe que o sintagma preposicional **[de futebol]** seja complemento de N, enquanto **[do Benfica]** seja um adjunto. Contudo, de acordo com as regras que apresentamos acima, a análise completa do sintagma nominal em (4.22) não consegue dar conta dessa distinção:

(4.22) [Um jogador de futebol do Benfica]. (Raposo, 1992: 204)

O que vemos acima é a repetição da regra (4.8), que diz que um **N'** pode expandir-se em um **N'** e um **PP**. Depois, temos a aplicação da regra (4.5), que autoriza que um **N'** possa expandir-se em um **N** e nada mais. Por isso, ambos os sintagmas preposicionais acima foram analisados como irmãos de N', ou seja, como estando em posição típica de adjuntos. Essa é uma falha nas regras que apresentamos.

O mesmo acontece com os adjetivos e com os advérbios: eles são tratados como elementos **intransitivos**. Em outras palavras, em sua subcategorização, esses elementos

não recebem complementos. Essa análise, contudo, não parece ser a mais correta. Sobre o sintagma nominal de (4.22), Raposo (1992: 204) propõe a seguinte análise, por exemplo:

(4.22) [Um jogador de futebol do Benfica]. (Raposo, 1992: 204)

Na representação acima, o sintagma preposicional **[de futebol]** é irmão de N **[jogador]** (ou seja, é **complemento** de N), enquanto o sintagma preposicional **[do Benfica]** é irmão de N' **[jogador de futebol]** (ou seja, modifica N'), estando, portanto, em posição de **adjunto** do NP. Se essa análise for a mais correta, precisaremos reelaborar as regras de descrição do sintagma nominal em relação à posição dos complementos.

E como sabemos que a análise de Raposo sobre o status dos PPs **[de futebol]** e **[do Benfica]** está correta? Quais serão os critérios que utilizaremos para distinguir substantivos (e adjetivos e advérbios e preposições e verbos) que aceitam (ou exigem) complemento dos que não aceitam complemento?

Aqui faremos uma observação bastante importante: não entraremos no árduo assunto sobre a distinção entre **núcleos transitivos** e **núcleos intransitivos**. Há diversos trabalhos interessantes que discutem a distinção entre núcleos lexicais transitivos e intransitivos. O assunto tem sido estudado tanto por gramáticos tradicionais já clássicos (como Bechara, 1960; e Luft, 1986), como por sintaticistas formais pioneiros, (como Jackendoff, 1977; e Williams, 1975). Para uma introdução ao problema, remetemos o leitor a Rodrigues (2000) e Carnie (2002); para uma proposta interessante para o PB, remetemos o leitor a Perini (1989, 2000).

De nosso ponto de vista, de quem está elaborando regras gramaticais de reescrita

para o PB sem lidar com os **traços gramaticais** dos itens lexicais, a diferença entre os adjuntos e complementos do sintagma será meramente a sua posição estrutural na árvore. Em outras palavras, iremos elaborar **macrorregras gramaticais** que nos permitem analisar estruturas sintáticas cujo núcleo exige ou aceita complementos e regras gramaticais que nos possibilitem analisar sintagmas cujo núcleo não aceita ou não exige complementos. Não iremos, portanto, analisar os itens léxicos em si em suas propriedades, por vezes idiossincráticas, de transitividade e seleção categorial. Daí entendermos nossas regras gramaticais como macrorregras de descrição gramatical.

Uma implementação computacional das regras gramaticais que iremos propor, ou um trabalho sintático mais fino, deve, evidentemente, levar em consideração a distinção entre núcleos transitivos e intransitivos, se quiser evitar a hipergeração de estruturas. Esse tipo de trabalho deve adotar uma **abordagem lexicalista**. A ideia central da abordagem lexicalista é justamente restringir as regras sintáticas com traços presentes nos itens lexicais. Esses traços são expressos através de variáveis que devem ser unificadas, percolando os nós em uma árvore sintática para garantir a boa formação da sentença, como fizemos em Othero (2004, 2006), por exemplo, e como mostramos no capítulo anterior.

Quando mostrarmos um elemento como complemento ou como adjunto em nossas análises (como fizemos com os NPs do exemplo de Raposo, acima), estaremos seguindo nossa própria intuição como linguista e falante nativo do português, além de análises já consagradas em teoria sintática do português. O foco de nosso trabalho é elaborar regras gramaticais (ou macrorregras gramaticais) que **prevejam estruturas com complementos e estruturas com adjuntos**, pois é fato consensual que as projeções máximas lexicais podem apresentar elementos que sejam subcategorizados pelos núcleos e elementos "acessórios" que apenas modificam o núcleo, como adjuntos[57].

Uma outra limitação nas regras que apresentamos anteriormente foi a inclusão de classes que não eram projeções máximas, como "det" e "pré-det", uma saída reconhecidamente *ad hoc*, que vai contra o princípio de maximalidade (cf. cap. 2) da teoria X-barra. Essa crítica nos foi feita por Alencar (2008: 9):

> Othero [2006] trata quantificadores em exemplos do tipo de (2) [todos, ambos] como pré-determinantes, inseridos por meio da regra NP → pré-det NP (p. 112). Essa análise viola um importante princípio do que seria uma teoria X-barra padrão: todo não-núcleo é um sintagma (Grewendorf,

[57] Sobre o assunto, cf. também Radford (1981, 1988) e Ernst (2001), além de Cançado (2003, 2005) para uma problematização sobre a distinção entre complementos e adjuntos de um ponto de vista semântico.

> Hamm & Sternefeld 1989: 202). Uma análise em termos da hipótese DP que obedecesse a esse princípio trataria todas, por exemplo, como QP na posição de especificador do DP. Othero, em vez disso, trata esse quantificador como adjunto, de modo que o analisador gramatical hipergera. Com efeito, a frase agramatical (11) é reconhecida como gramatical pelo Grammar Play:
> (11) *Todas ambas ambas todas ambas as meninas dormem tranqüilamente.

Para superar limitações como essas, reformulamos as regras para a descrição do sintagma nominal. Na verdade, o sintagma nominal como apresentamos nas regras (4.1) a (4.8) deu lugar às regras do DP, de (4.23) a (4.27):

(4.23) DP → D'
(4.24) D' → D
(4.25) D' → D NP
(4.26) D' → D PossP
(4.27) D' → D NumP

Além dessas regras, elaboramos as seguintes regras gramaticais para o **NP**, o **NumP**, o **PossP** e o **QP**:

(4.28) NP → N'
(4.29) N' → N
(4.30) N' → N' AP
(4.31) N' → AP N'
(4.32) N' → N' PP
(4.33) N' → N PP

(4.34) NumP → Num'
(4.35) Num' → Num NP
(4.36) Num' → Num PP

(4.37) PossP → Poss'
(4.38) Poss' → Poss NP
(4.39) Poss' → NP Poss
(4.40) Poss' → Poss NumP

(4.41) QP → Q'

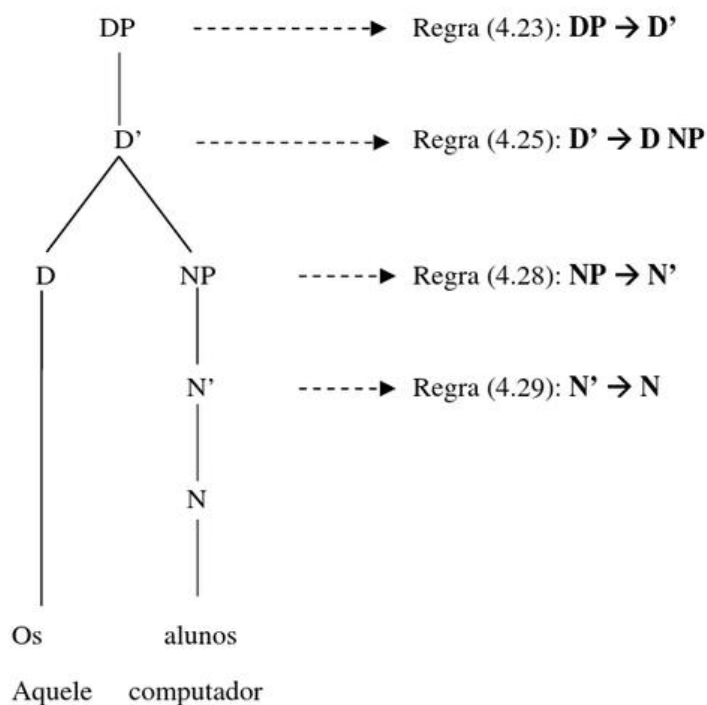
(4.42) Q' → Q DP

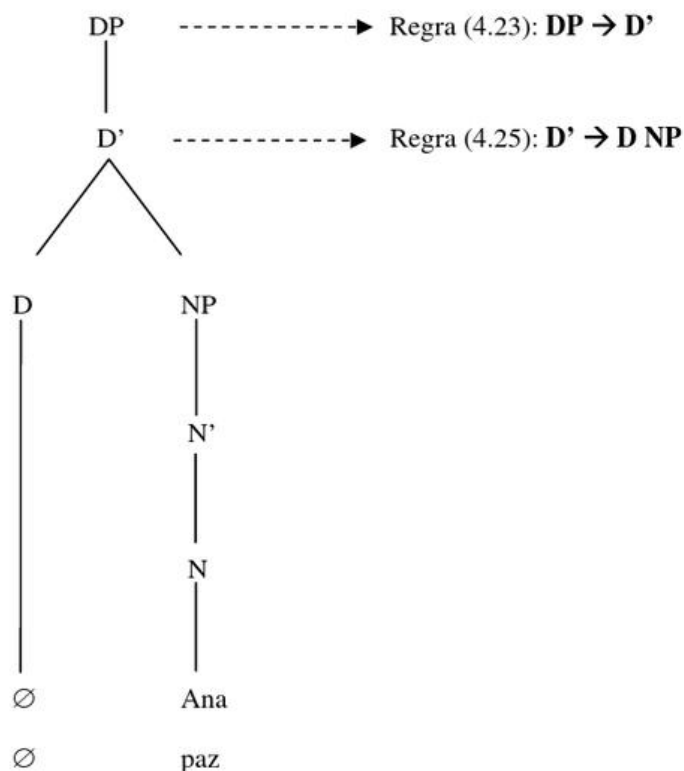
(4.43) Q' → Q PP

(4.44) Q' → DP Q

Vejamos a análise de alguns exemplos que motivaram nossas regras para discutir nossa proposta de descrição do DP.

(4.45) [Os alunos] adoraram o livro.

(4.46) [Aquele computador] é melhor do que este.

(4.47) [Ana] chegou.

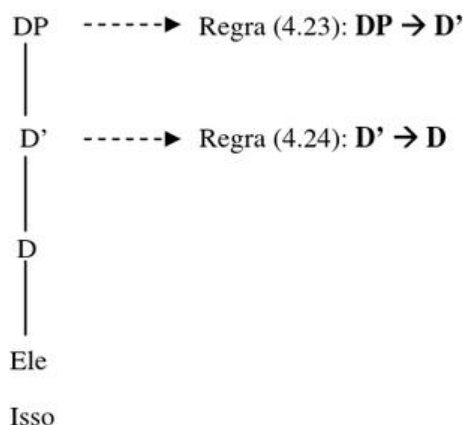
(4.48) Todos desejam [paz].

As estruturas arbóreas acima são talvez as mais consensuais na descrição do DP em português. A regra (4.25) permitiu a formação tanto dos DPs em (4.45) e (4.46) quanto dos DPs em (4.47) e (4.48), com o elemento D nulo (aqui representado pelo símbolo ∅). Há fortes indícios de que o D possa ser expresso como elemento foneticamente não preenchido nas línguas[58]. Por isso, adotamos essa análise. Porém, sabemos que a implementação computacional de elementos vazios pode representar um problema[59].

Vejamos a motivação para a regra (4.24).

(4.49) [Ele] comprou um livro interessante.

(4.50) [Isso] parece interessante.

[58] Remetemos o leitor a Giusti (1997), Vangsnes (2001), Pereltsvaig (2005) e algumas das referências citadas ali para discussão de alguns dos argumentos principais.

[59] No entanto, a proposta de Black (1997), que utiliza o formalismo PATR, consegue lidar com alguns tipos de movimento e vestígios na análise das sentenças em inglês. Cf. também algumas propostas de *parsing* em Prolog que adotam modelos da GB, como Kuhns (1986, 1990) e Chen, Lin & Wu (1988).

O núcleo do DP em português pode ser um **artigo**, um **elemento vazio**, um **pronome demonstrativo** ou um **pronome pessoal**. Os determinantes, assim como demais núcleos funcionais, subcategorizam o tipo sintático de seu complemento, ou seja, **c-selecionam** seus complementos. Por isso, encontramos dois tipos de núcleos D: aqueles que selecionam complementos (os complementos serão NPs, PossPs e NumPs, como veremos) e aqueles que são intransitivos, ou seja, que não selecionam complementos.

Em português, os itens lexicais que podem ocupar essa última posição, a de D intransitivo, são os **pronomes pessoais** e os **pronomes demonstrativos**, como mostramos nos exemplos (4.49) e (4.50). E os itens lexicais que podem ocupar a posição de D transitivos são os **artigos**, os **pronomes demonstrativos** (com exceção de **isso** e **aquilo**) e os **elementos-QU interrogativos**, como em **[Que livro] o João está lendo?**[60].

Esses elementos estão em distribuição complementar na estrutura dos agrupamentos nominais em português e, por isso, são de mesma natureza. Não podemos ter as sequências abaixo, em português:

(4.51) *[O ele] chegou.
(4.52) *[Ele isso] chegou.
(4.53) *[A esta menina] saiu.
(4.54) *[Que este livro] João está lendo?
(4.55) *[Uma esta menina] gosta de Sintaxe.

---

[60] Alguns linguistas argumentam que os pronomes pessoais também podem ocupar a posição de núcleo D transitivos, como no caso dos DPs [*nós linguistas*], [*vocês meninas*], etc. Esses pronomes pessoais poderiam receber o traço **[+transitivo]**. Ignoraremos o fato aqui, pois não nos parece que o NP que segue o pronome esteja atuando como seu complemento. A esse respeito, cf. Radford (2007b), em análise do inglês, e David (2006a,b, 2007), em análise do português.

Etc.

Interessante que o mesmo não acontece na distribuição e no comportamento sintático de outros itens que costumamos encontrar nos agrupamentos nominais em PB, como os **pronomes possessivos** e os **numerais**, como vemos nas seguintes construções:

(4.56)

a. [Meus sobrinhos]

b. [Os meus sobrinhos]

c. [Estes meus sobrinhos]

d. [Estes meus dois sobrinhos]

(4.57)

a. João engordou [dois porcos].

b. João engordou [os dois porcos].

c. João engordou [estes dois porcos].

d. João engordou [seus dois porcos].

e. João engordou [quais dois porcos]?

Os pronomes possessivos e os numerais costumam ser classificados como determinantes (ou como um tipo de determinantes) em PB[61]. No entanto, não podemos classificá-los como **determinantes**. Afinal, como vimos nos exemplos acima, eles não estão em distribuição complementar com os determinantes que havíamos estudado. E não podemos considerá-los adjetivos porque seu comportamento gramatical é distinto do comportamento dos adjetivos. Enquanto os adjetivos são núcleos lexicais, com conteúdo semântico (*bonito, rápido, grande*), os possessivos são núcleos funcionais que têm um papel gramatical dentro do agrupamento nominal[62]. Além do mais, os adjetivos podem ser utilizados de maneira recursiva para modificar um N (*querido amigo / querido prezado amigo / querido prezado excelentíssimo amigo*, etc.), enquanto os possessivos não têm

[61] Souza e Silva & Koch (1993), Bick (2000), Othero (2004, 2006). Há também quem considere os possessivos como uma subclasse dos adjetivos, por apresentarem comportamento sintático semelhante aos adjetivos na estrutura do DP, como Gonzaga (2003, 2006).

[62] Essa distinção "semântica" entre elemento **lexical** e elemento **funcional** remonta à tradição de gramáticos descritivos como Bolinger & Sears (1968), que utilizavam a dicotomia **palavras de conteúdo** (*contentives*) vs. **palavras de função** (*functors*). As primeiras têm conteúdo descritivo, semântico; as segundas trazem basicamente informações de caráter gramatical. Parece-nos justamente que os pronomes possessivos são "palavras de função".

essa característica (*meu amigo / *meu teu amigo / *meu teu nosso amigo*, etc.). Uma outra evidência diz respeito à ordem de modificação dos adjetivos e dos possessivos dentro do DP (assunto que retomaremos com mais detalhes adiante, ainda neste capítulo): se os dois elementos encontram-se modificando um N em posição pré-nuclear, o adjetivo deve necessariamente estar mais próximo ao nome: *meu querido amigo / *querido meu amigo*.

É comum encontrarmos os pronomes possessivos classificados como **D** em inglês (Radford, 1997a, b, por exemplo), já que, nessa língua, eles se encontram em distribuição complementar com os artigos e com os pronomes demonstrativos. Sua posição dentro do sintagma é intercambiável, e uma sequência com os dois elementos é agramatical, como mostramos abaixo:

(4.58) [My dog] is there.
(4.59) [The dog] is there.
(4.60) *[The my dog] is there.

Em português, no entanto, todas essas sequências são gramaticais: *meu cachorro / o cachorro / o meu cachorro está lá*.

Uma outra saída comum adotada nos estudos da sintaxe do sintagma nominal em PB é classificar esses elementos como uma **subclasse dos determinantes**, ou seja, como pré- ou pós-determinantes. Foi a saída que adotamos em Othero (2004 e 2006), por exemplo, e que apresentamos no capítulo 3. Essa "solução" também não é boa. Ela apresenta, pelo menos, duas restrições graves, uma de ordem empírica e outra de natureza teórica.

Vejamos as seguintes sequências em português:

(4.61) [O computador]
(4.62) [O meu computador]
(4.63) [Os dois computadores]

O pronome possessivo aparece em posição pós-determinante em (4.62), assim como o numeral em (4.63). À primeira vista, poderíamos pensar em classificar esses dois elementos como pós-determinantes. No entanto, os dados abaixo nos mostram que esses elementos não são da mesma natureza, pois não estão em distribuição complementar:

(4.64) [Os meus dois computadores]

(4.65) *[Os dois meus computadores]

(4.66) *[Meus os dois computadores]

(4.67) *[Dois os meus computadores]

(4.68) *[Meus dois os computadores]

Os dois elementos podem co-ocorrer no agrupamento nominal, mas há uma ordem linear que deve ser respeitada entre eles. O pronome possessivo deve vir antes do numeral, e ambos devem figurar à direita do determinante. Isso parece complicar a classificação desses elementos dentro do NP.

Além disso, de um ponto de vista interno à teoria, se quisermos ser fiéis aos princípios teóricos que norteiam a descrição gramatical do modelo proposto pela teoria X-barra, não podemos postular elementos "soltos", que não sejam projeções máximas, e que estejam desrespeitando, por exemplo, os princípios da lexicalidade e da maximalidade. Essa crítica, como já vimos, nos foi feita por Alencar (2008).

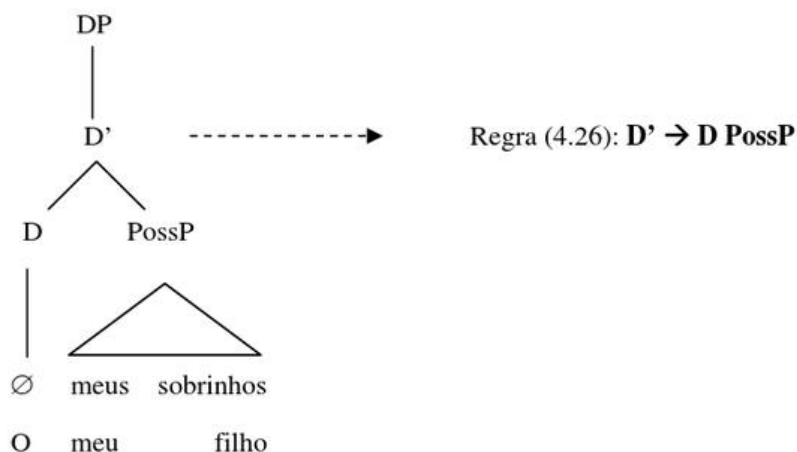
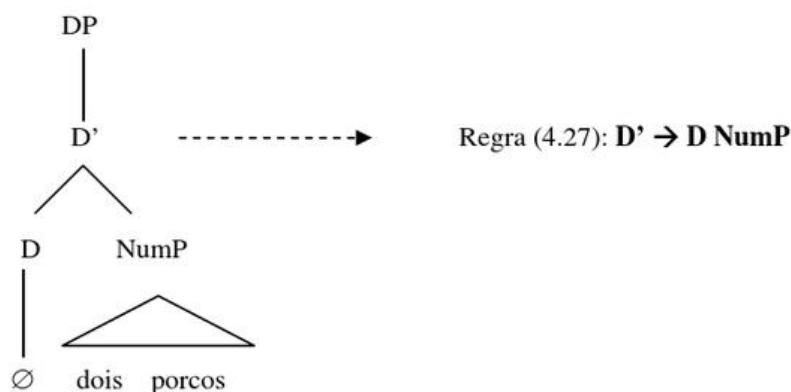
Para superar essas limitações, propusemos dois tipos de sintagma em PB que figuram dentro da estrutura do DP: o sintagma numeral (**NumP**) e o sintagma possessivo (**PossP**)[63]. Apresentaremos nossas análises de DPs em português que parecem mostrar que nossa proposta é adequada. Além de estudarmos com cuidado a estrutura interna dos agrupamentos nominais em PB para propor esses dois núcleos funcionais, nos beneficiamos da análise de Perini (2000), que também reconhece que os pronomes possessivos e os numerais não podem ser classificados como determinantes em português.

Vejamos algumas análises do DP com esses dois núcleos funcionais. Mais tarde, mostraremos a estrutura interna dos PossPs e dos NumPs.

(4.69) [Meus sobrinhos] acabam de sair.

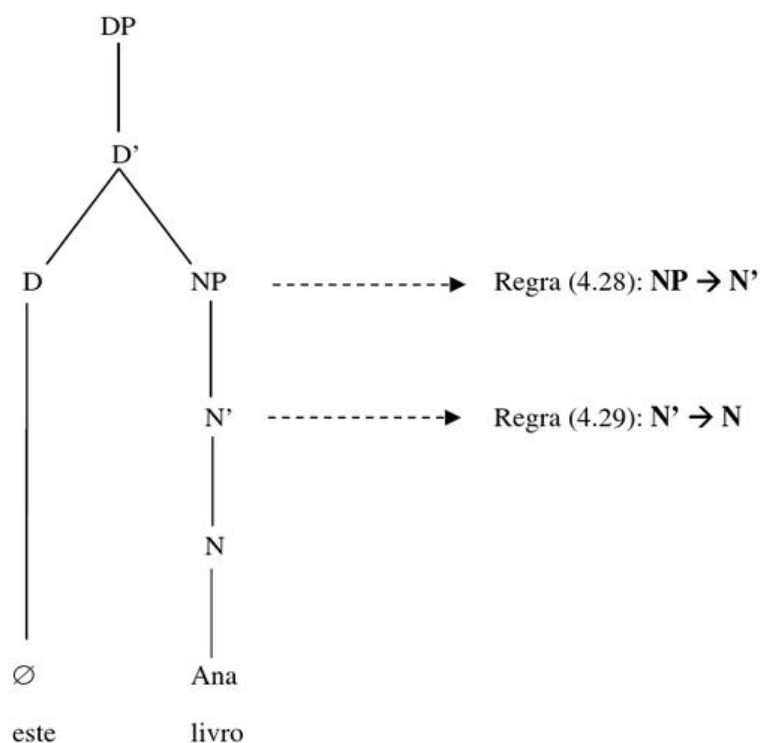
(4.70) [O meu filho] acaba de chegar.

---

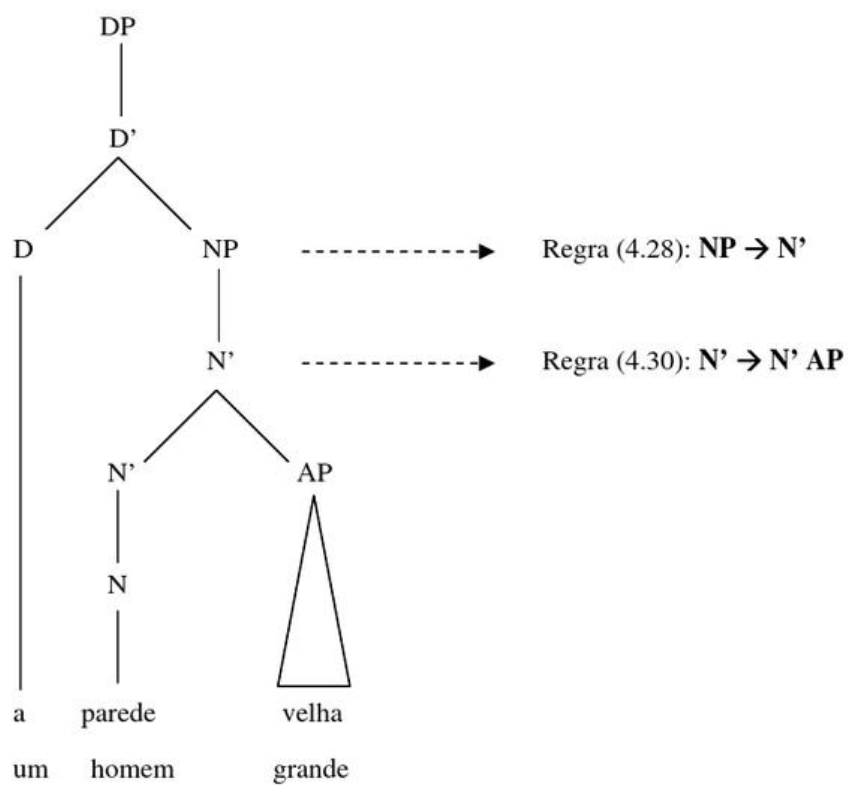
[63] Sabemos que postular núcleos funcionais para descrever estruturas gramaticais na língua não é um recurso visto com bons olhos por alguns sintaticistas. Por isso, tivemos muito cuidado ao postular o NumP e o PossP. Realmente acreditamos que essas projeções podem esclarecer alguns pontos sobre a estrutura interna do DP em português. Recentemente (em agosto de 2008, durante o *VII Workshop on Formal Linguistics*, em Curitiba), ouvimos o seguinte comentário de um colega sintaticista: "não parece econômico ver que cada linguista postule um núcleo funcional na descrição da estrutura da sentença de uma dada língua. Mas, afinal de contas, quem garante que a estrutura gramatical de uma língua seja mesmo econômica? Se realmente houver 32 projeções dominando um sintagma verbal, temos de encarar isso como um fato da língua e não nos prender à economia que a teoria prega. Fazer o quê?".

(4.71) João engordou [dois porcos].

As regras do DP preveem estruturas com o núcleo D transitivo, que aceitam um complemento (NP, PossP ou NumP), e com um núcleo D intransitivo, que não necessita de complementos. Como dissemos, não estudaremos os traços lexicais que compõem cada elemento lexical. Mas podemos pensar em um traço **[±transitivo]** que poderia ser implementado na descrição do núcleo D. Os núcleos D transitivos c-selecionam como complemento um NP (regra 4.25), um PossP (regra 4.26) ou um NumP (regra 4.27).

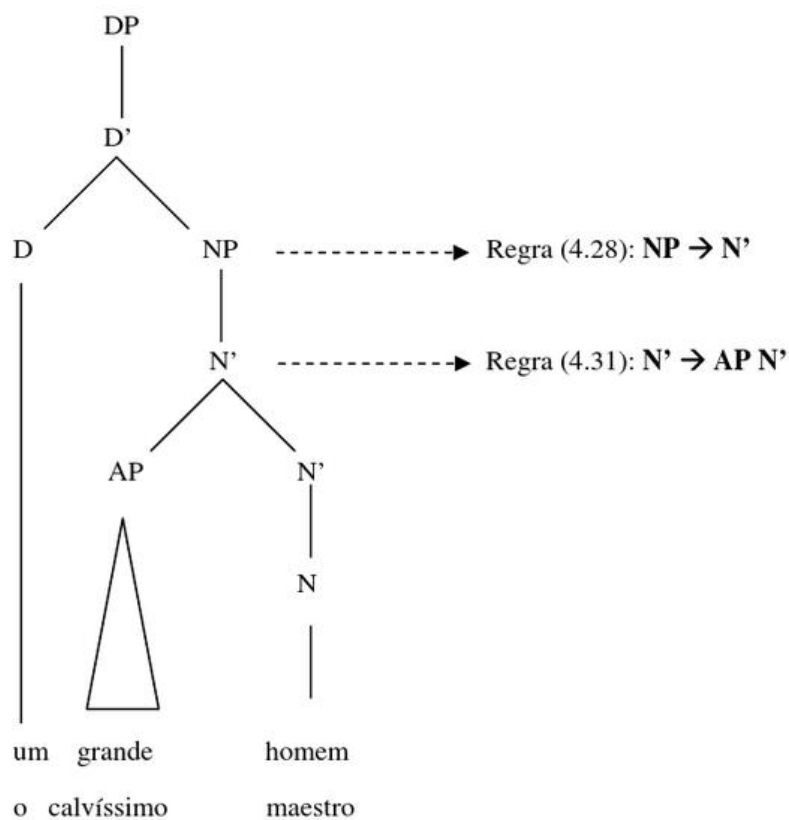
Antes de analisarmos a estrutura interna do PossP e do NumP, passemos às regras do NP. Uma das limitações das regras que propusemos em Othero (2004), como vimos, foi a indistinção entre os complementos e adjuntos na estrutura interna do NP. Por isso, reelaboramos as regras para que elas pudessem mostrar essa diferença. Comecemos com análises de NPs intransitivos, que não apresentam complementos, como os previstos pelas regras (4.29) a (4.32).

(4.72) [Ana] chegou.

(4.73) [Este livro] parece interessante.

A estrutura arbórea acima apresenta um NP sem qualquer complemento ou adjunto. Mas sabemos que o NP pode apresentar complementos e adjuntos. Comecemos por ver as regras que elaboramos para analisar NPs com adjuntos (os adjuntos do NP podem ser um **sintagma adjetival** (AP) ou **preposicional** (PP)).

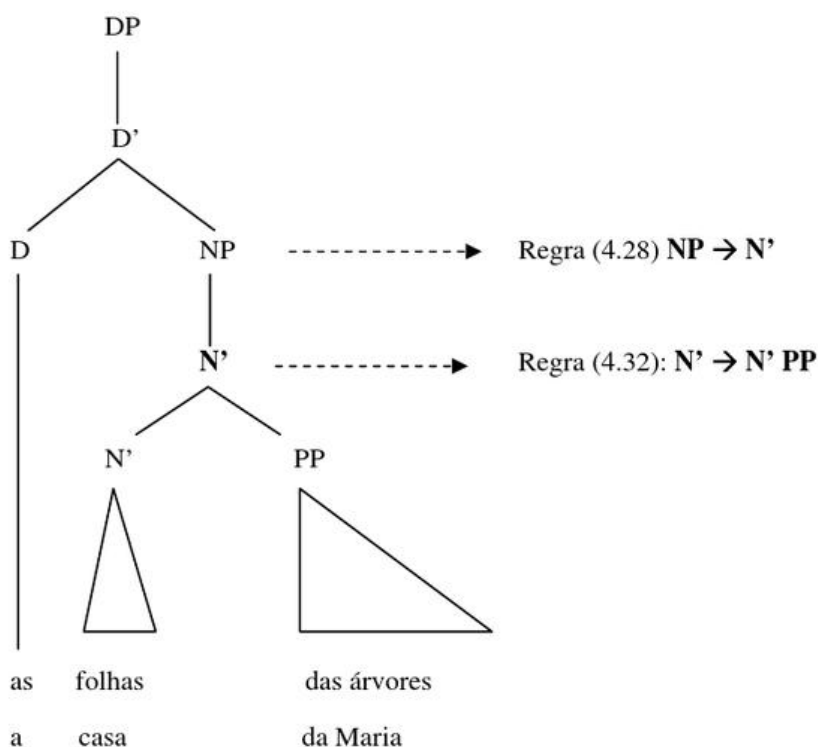
(4.74) Eu pintei [a parede velha].

(4.75) Oscar é [um homem grande].

(4.76) Oscar é [um grande homem].

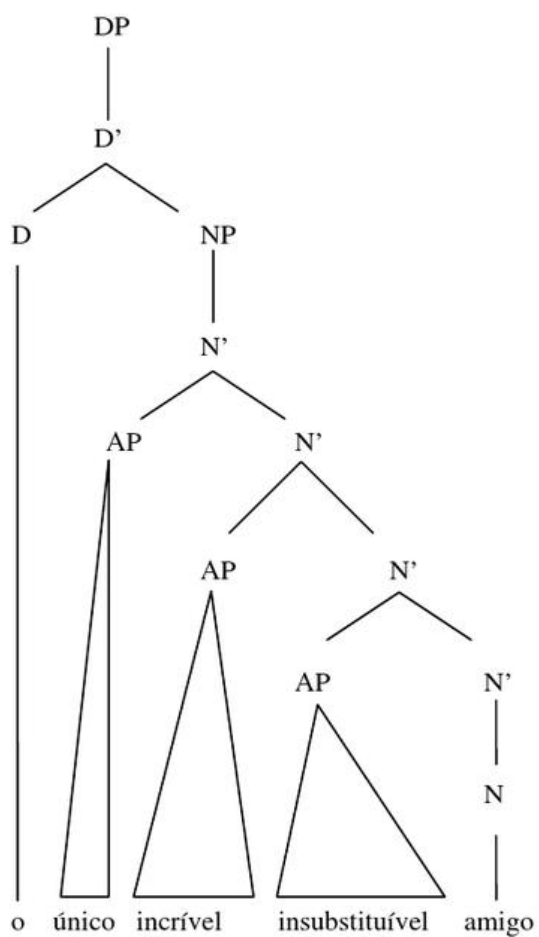
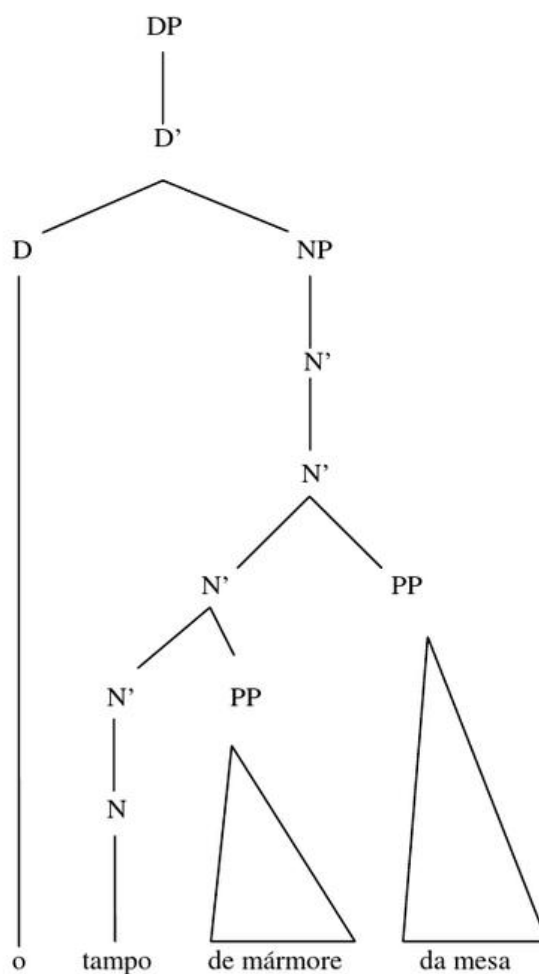
(4.77) [O calvíssimo maestro] acaba de chegar.

Em termos estritamente sintáticos e sintaticamente estritos, há pouca diferença entre as regras (4.30) e (4.31). A regra (4.30) prevê um AP como adjunto **à direita** do N', enquanto a regra (4.31) prevê um AP adjunto **à esquerda** do N'. No entanto, a modificação adjetival à esquerda ou à direita do N' envolve alguns aspectos semânticos interessantes (*um grande homem* não recebe a mesma interpretação que *um homem grande*, por exemplo)[64].

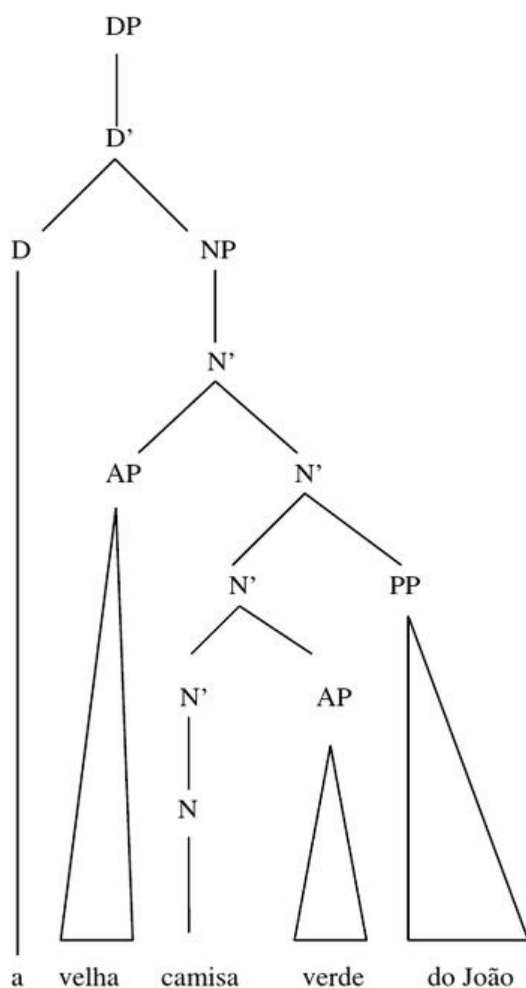
Ainda tratando da modificação nominal, vemos a regra (4.32), que nos permite analisar NPs com modificadores preposicionais em posição de adjuntos.

(4.78) [As folhas das árvores] começaram a cair.

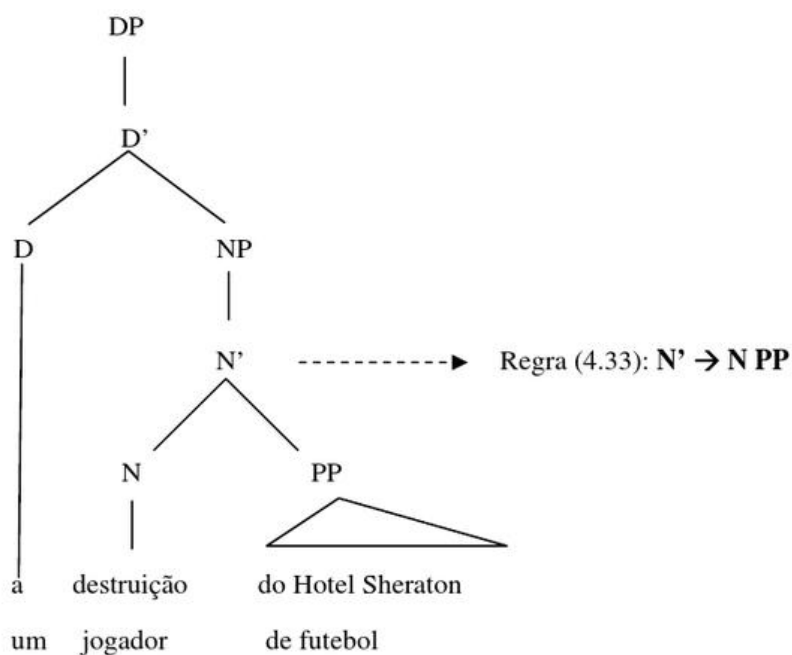
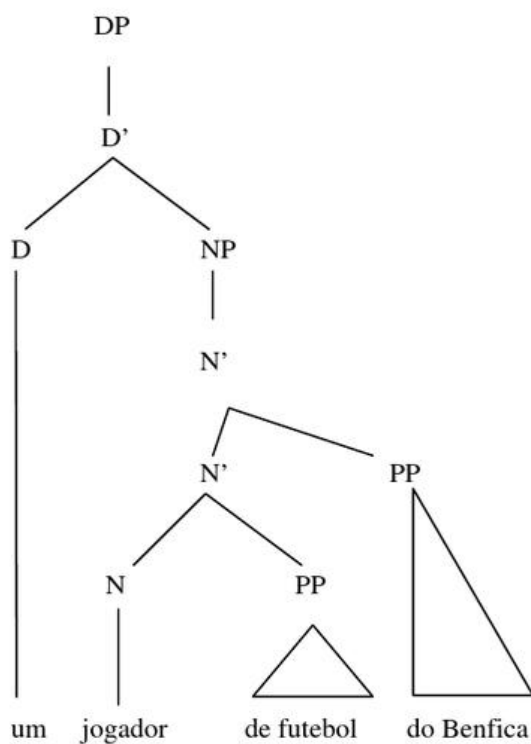
(4.79) João entrou [na casa da Maria].

[64] Não lidaremos com o assunto aqui, já que essa distinção semântica não se reflete na estrutura sintática que estamos analisando além da distinção de ordem linear que já apontamos. Sobre o tema, cf. Bernstein (1991), Giorgi & Longobardi (1991), Valois (1991), além de dois trabalhos interessantes que lidam com dados do português: Menuzzi (1992), cuja proposta é descrever o comportamento e a interpretação de adjetivos pré- e pós-nominais em PB dentro da teoria gerativa; e Callou et al. (2003), que estudam a distribuição dos adjetivos dentro do NP em PB e PE, com base em corpora das duas línguas – os autores chegam à conclusão de que "no estudo da posição do Adjectivo no Sintagma Nominal intervêm factores muito diversos: morfossintácticos e sintácticos, de compatibilidade semântica e pragmática entre os adjectivos e os nomes, factores estilísticos e rítmicos, entre outros".

Repare que essas regras são recursivas. Isso quer dizer que elas permitem a repetição de APs e PPs adjuntos ao N’, como nas sentenças seguintes:

(4.80) [O único incrível insubstituível amigo] acaba de chegar.

(4.81) [O tampo de mármore da mesa] rachou.

Os exemplos acima apresentam mais de um sintagma (adjetival ou preposicional) modificando o NP, o que é previsível pelas regras de nossa gramática. Além disso, as regras, como estão até agora, não permitem a formação de sintagmas agramaticais, com modificação de um PP à esquerda do N', como as sequências **o de mármore tampo* ou **as das árvores folhas*. E a modificação múltipla de um N'por APs e PPs também é prevista, como vemos na análise abaixo:

(4.82) [A velha camisa verde do João].

Essas análises demonstram que as regras gramaticais que formulamos para a descrição do DP e do NP em português parecem estar adequadas, prevendo a boa formação de sintagmas gramaticais e barrando alguns sintagmas agramaticais em PB. Passemos, por fim, à análise da regra (4.33), que prevê um N transitivo com complemento preposicional.

(4.83) Eu presenciei [a destruição do Hotel Sheraton].

(4.22) [Um jogador de futebol] do Benfica. (Raposo, 1992: 204)

Analisando o DP completo **[Um jogador de futebol do Benfica]**, percebemos a distinção que as regras fazem entre complementos (irmãos do núcleo) e adjuntos (irmãos da projeção intermediária X'):

Na estrutura arbórea acima, o PP **[de futebol]** é complemento do N – daí estar em

relação de irmandade com N –, enquanto o PP **[do Benfica]** é adjunto do NP – daí estar em relação de irmandade com o N' e não com o N.

Passemos às regras gramaticais (4.34) a (4.44), que tratam da descrição das projeções máximas funcionais que postulamos para a descrição do DP em português. Repetimos as regras aqui por conveniência:

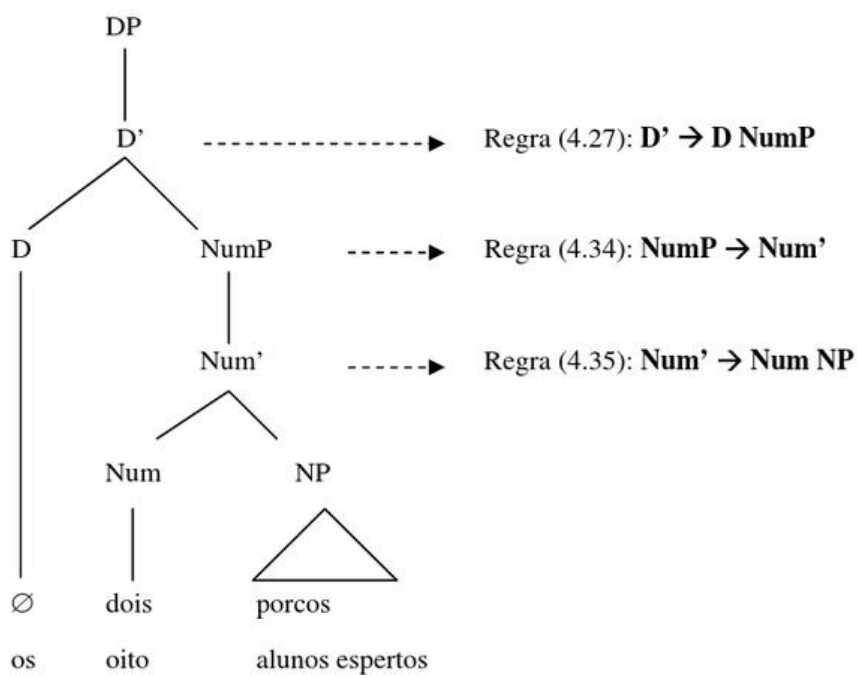
(4.34) NumP → Num'
(4.35) Num' → Num NP
(4.36) Num' → Num PP

(4.37) PossP → Poss'
(4.38) Poss' → Poss NP
(4.39) Poss' → NP Poss
(4.40) Poss' → Poss NumP

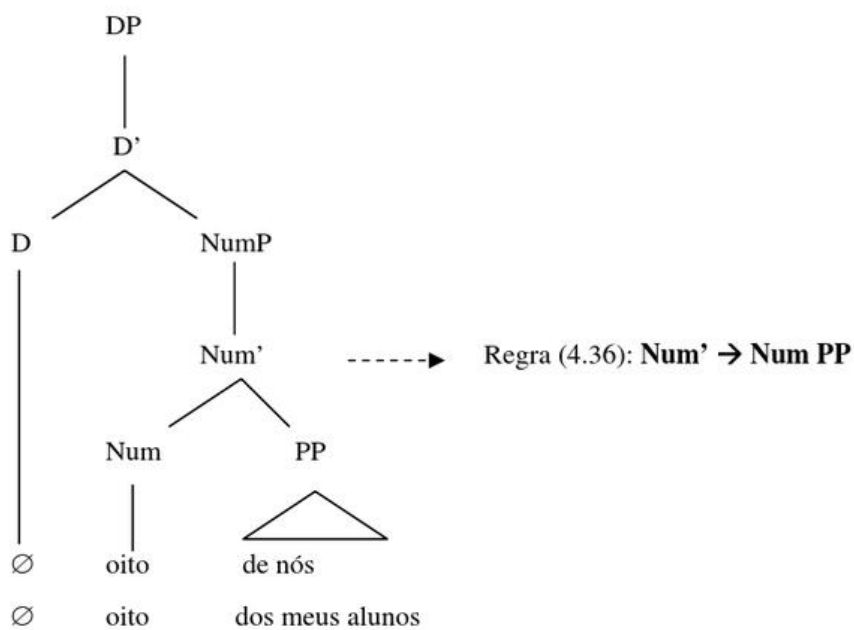
(4.41) QP → Q'
(4.42) Q' → Q DP
(4.43) Q' → Q PP
(4.44) Q' → DP Q

Postulamos a existência dessas categorias na estrutura do DP para dar conta de determinados agrupamentos nominais de maneira mais elegante e eficiente do que havíamos proposto em trabalhos anteriores. **Elegante** porque essas regras estão mais afinadas com a teoria gramatical da X-barra; **eficientes** porque acreditamos que, com essas regras, conseguimos analisar corretamente a estrutura interna do DP, lidando com o problema da ordenação de elementos como determinantes, pronomes possessivos e numerais no DP e NP.

Vejamos a motivação para as regras que apresentamos acima. Comecemos pelo núcleo **Num**, expresso por um numeral. Vejamos a análise dos seguintes DPs:

(4.84) João engordou [dois porcos].
(4.85) [Os oito alunos espertos].

Repare que o **NumP** domina o **NP**. Graças a essa relação de dominância, conseguimos analisar com sucesso sintagmas como os seguintes:

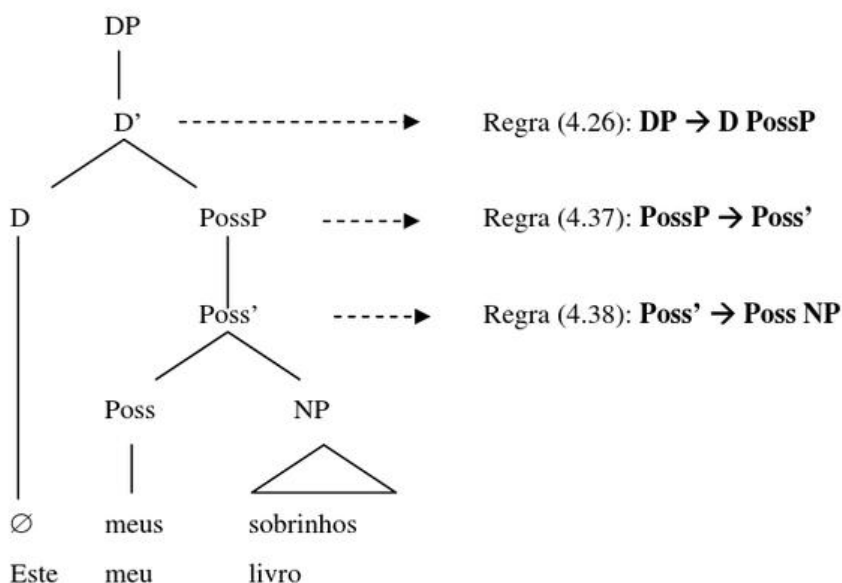
(4.86) [Oito de nós].

(4.87) [Oito dos meus alunos].

Uma vez que o sintagma máximo **NumP** contém o NP (e não está contido por ele),

podemos também estipular que NumP domine um PP, como mostramos na análise acima, em (4.86) e (4.87). De maneira correspondente, postulando que o DP domina o NumP, podemos também analisar sem problemas os sintagmas de (4.84) e (4.85), que apresentam um determinante foneticamente preenchido e um determinante vazio[65]. Além disso, essa análise será crucial em casos de sintagmas com pronomes possessivos: as regras do NumP irão interagir de maneira bastante harmônica com as regras referentes à estrutura do PossP, como veremos a seguir. Finalmente, uma outra vantagem dessa análise é que não é necessário postular que sequências como [o aluno], [meu aluno], [oito alunos], [alunos], etc. sejam de natureza diferente: todos são DPs e, portanto, sendo da mesma categoria, todos devem apresentar o mesmo comportamento sintático.

Passemos às análise de DPs contendo elementos possessivos.

(4.88) [Meus sobrinhos] são umas pestes.

(4.89) [Este meu livro] é muito interessante.

O DP deve dominar os núcleos funcionais NumP e PossP. Aqui, com o PossP, vemos mais um argumento para isso: se aceitarmos uma estrutura $_{PossP}[\ _{DP}[\ ...\ ]\ ]\ ]$, ou seja, com um PossP dominando um DP, estaríamos autorizando formações claramente agramaticais na língua, como **meus os sobrinhos*, **este meu este livro* e mesmo **meus*

[65] A ordem dos elementos do DP em português reflete a ordem hierárquica universal dos elementos dentro do agrupamento nominal proposta por Cinque (2005): **D > Num > Adj > N** (ainda que, como vimos, a ordem do adjetivo seja relativamente menos rígida, cf. Menuzzi, 1992, e Perini, 2000).

meus sobrinhos ou **este meu meu livro*. As regras acima preveem a agramaticalidade dessas sequências corretamente. Ao autorizarmos somente estruturas do tipo $_{DP}$[$_{PossP}$[$_{NP}$[...]]], evitamos esse tipo de má formação do sintagma.

Contudo, ainda não conseguimos lidar adequadamente com os DPs abaixo:

(4.90) A Maria leu [**o meu** livro].
(4.91) ?? A Maria leu [**o** livro **meu**][66].
(4.92) ?? A Maria leu [**um meu** livro].
(4.93) A Maria leu [**um** livro **meu**].

O que acontece aqui é que os pronomes possessivos se comportam de maneira distinta quando na presença de artigos definidos, como em (4.90) e (4.91), ou artigos indefinidos, como em (4.92) e (4.93). Isso sugere que o núcleo D deve conter um traço **[± definido]** que afetará a posição do pronome possessivo dentro do DP[67]. Em relação à sua posição no DP, encontramos o seguinte:

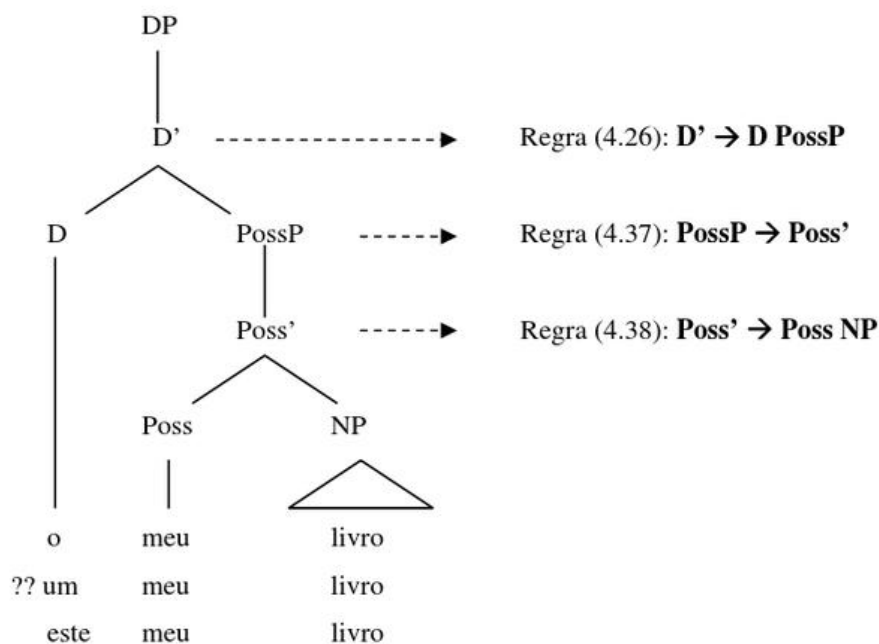
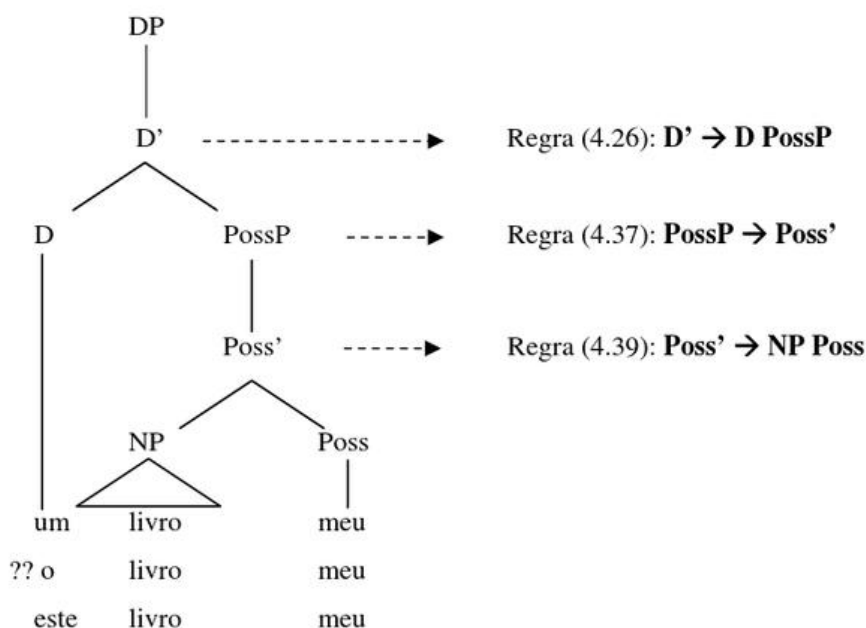
- com o artigo definido, o pronome possessivo deve ocupar posição pré-nominal (ver, no entanto, observação da nota 66);
- com o artigo indefinido, o pronome possessivo deve ocupar posição pós-nominal.

E quando temos um pronome demonstrativo, o possessivo pode ocupar qualquer uma das duas posições:

(4.94) A Maria leu [este meu livro].
(4.95) A Maria leu [este livro meu].

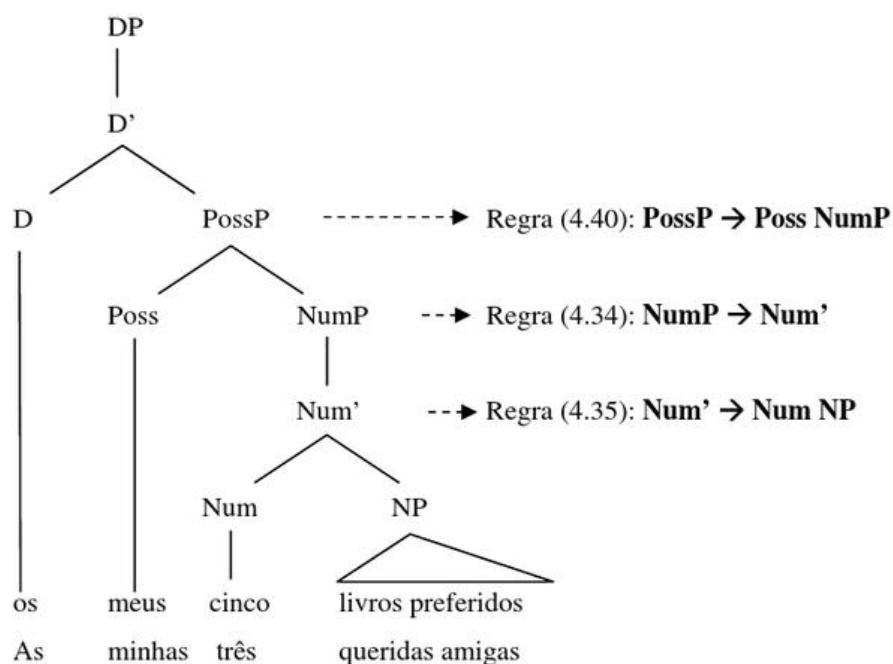
[66] Magdiel M. Aragão Neto, em comunicação pessoal, não julga essa sentença estranha, uma vez que, "na poesia, por exemplo, é muito comum essa estrutura e ninguém estranha". Além do mais, encontramos esse tipo de ordem em PB, ainda que de maneira restrita, como neste trecho da oração do Pai Nosso: "**O pão nosso** de cada dia nos dai hoje (...)". Essa estrutura era comum em português arcaico (Mattos e Silva, 2006) e ainda encontramos algumas ocorrências no português falado no Brasil (Moura Neves, 1993). Remetemos o leitor a Castro (2006a, b) e Brito (2007), que comparam o comportamento desse tipo de construção em PE e PB e trazem alternativas de análise para o fenômeno, nas duas variedades do português.

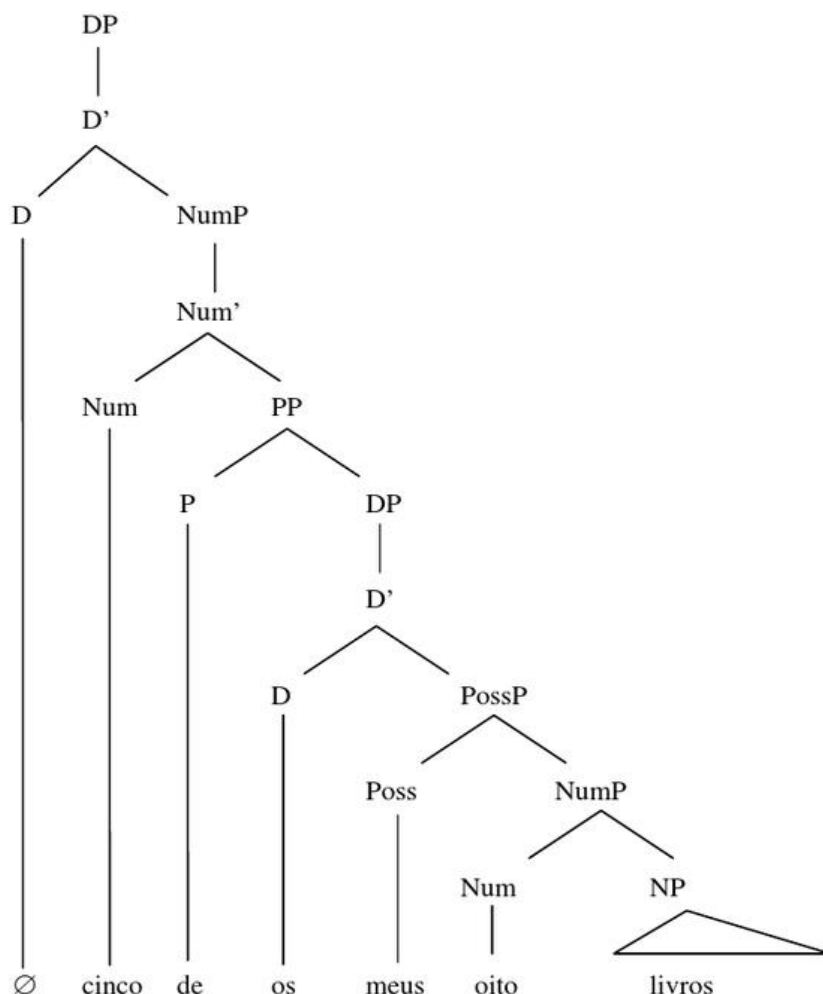
[67] A existência de um traço [± definido] é sugerida por Martinho (1998) e David (2007), para o português europeu e brasileiro, respectivamente. A mesma intuição também é compartilhada por Gonzaga (2003) ao lidar com o problema.

Isso não representa de fato um problema para nossas (**macro**)regras gramaticais, uma vez que nossas regras permitem a formação de estruturas bem formadas, como mostramos nas representações arbóreas abaixo. O problema, na verdade, é que não conseguimos restringir a formação de estruturas mal formadas na língua. Ou seja, a mesma regra que garante e prevê a boa formação do DP em (4.90) e (4.94) também garante a boa formação do estranho DP em (4.92).

O mesmo ocorre quando da descrição dos DPs (4.93) e (4.95): as regras preveem sua boa formação, mas não excluem o DP estranho de (4.92):

Como sugerimos acima, a restrição que irá permitir a boa formação de (4.90) e (4.93) e excluir (4.91) e (4.92) pode estar relacionada à implementação de um traço lexical [± **definido**]. Isso significa que as regras que estamos propondo parecem estar bem formuladas. Elas estão, na verdade, à espera da interação com o componente lexical, que irá trazer as informações necessárias sobre as categorias morfossintáticas e seu relacionamento dentro do DP, algo como propusemos na implementação da gramática do Grammar Play (cf. cap. 3), ou como fazem algumas abordagens gramaticais baseadas na unificação ou checagem de traços.

Passemos à última regra envolvendo os sintagmas PossP e NumP: a regra (4.40). Ela irá descrever estruturas de DPs como os seguintes:

(4.96) A Maria leu [os meus cinco livros preferidos].

(4.97) [As minhas três queridas amigas] acabam de chegar.

A interação entre as regras dos elementos **Num** e **Poss** permitirá analisarmos corretamente sintagmas como (4.98), com um NumP dominando um PossP dominando um NumP, de maneira recursiva:

(4.98) [Cinco dos meus oito livros] desapareceram[68].

Finalmente, passemos às análises envolvendo a projeção máxima QP. Adotamos o QP seguindo outras propostas e análises envolvendo sintagmas com quantificadores, como Shlonsky (1991), Raposo (1992), Radford (1997a,b), Vicente (2006), entre outros. Há na literatura diversos argumentos em favor da hipótese de se postular um núcleo funcional QP. Apresentaremos dois problemas de natureza sintática que parecem se beneficiar da postulação de um sintagma quantificador para análises em português.

O primeiro diz respeito a uma restrição em PB, conhecida como a **restrição do pronome pleno**. Esse fenômeno só acontece em línguas pro-drop (como o português)[69].

[68] No entanto, sempre que lidamos com recursão, estamos lidando com um mecanismo extremamente poderoso das línguas e – até hoje – de difícil tratamento computacional. O problema aqui é que essas regras supergeram estruturas. Não há nada que impeça a formação do sintagma claramente inaceitável **[cinco de meus oito de meus dez de meus vinte livros]**, por exemplo.
[69] Cf. Montalbetti (1984), Rothman & Iverson (2007), Molsing (2008).

Em línguas pro-drop, um pronome pleno não pode servir como elemento anafórico de um QP, como vemos no contraste das sentenças em português e inglês (língua não-pro-drop) abaixo:

(4.99) [Todos os alunos]$_i$ disseram que eles$_{*i/j}$ são inocentes.

(4.100) [All the students]$_i$ said they$_{i/j}$ are innocent.

Em PB, é impossível ligar um pronome pleno a um QP. No entanto, essa restrição desaparece quando o referente é um simples DP:

(4.101) [Os alunos]$_i$ disseram que eles$_{i/j}$ são inocentes.

Postulando a distinção entre um QP e um DP, percebemos que a restrição do pronome pleno se dá, em PB, apenas com QPs. A restrição não atua sobre DPs, apontando para o fato de que eles são efetivamente de natureza distinta[70].

Um outro argumento em favor da ideia do QP em português vem da análise de Costa (1997, 1998), quando ele mostra algumas diferenças no comportamento de sujeitos pré-verbais definidos e indefinidos. Para Costa (1998: 115), "diferentemente dos sujeitos pré-verbais definidos, os sujeitos pré-verbais indefinidos não podem co-ocorrer com extração de elementos-wh", por motivos de restrições no movimento A-barra[71]. Como exemplos, temos os seguintes pares:

(4.102)

a. Tu sabes quais livros a mulher leu __?

b. *Tu sabes quais livros uma mulher leu __?

(4.103)

a. Que filme o menino alugou __?

---

[70] Lembre-se de que o mesmo servirá para DPs contendo NumPs e PossPs, como (i) e (ii). Vimos que essas projeções estão contidas no DP; logo, são DPs. Repare como (i) e (ii) são gramaticais:
i. [Dois alunos]$_i$ disseram que eles$_{i/j}$ são inocentes.
ii. [Meus alunos]$_i$ disseram que eles$_{i/j}$ são inocentes.
Essa é mais uma evidência em favor de nossas regras gramaticais, que postulam que os NumPs e os PossPs são, de fato, DPs e que os agrupamentos com quantificadores são de outra natureza.

[71] Cf. argumentação em Costa (1997, 1998). Trecho original: "Differently from preverbal definite subjects, preverbal indefinite subjects may not cooccur with wh-extraction". Sobre movimento A-barra, cf. capítulo 6 de Culicover (1997), por exemplo.

b. *Que filme um menino alugou__?

O julgamento de gramaticalidade que distingue sentenças com movimento-wh com sujeitos pré-verbais definidos e indefinidos parece corroborar o argumento de Costa. No entanto, ele percebe que os julgamentos são afetados quando os sujeitos são DPs com elementos como *vários* (cf. Costa, 1998: 116):

(4.104) ?/?? Quais livros várias pessoas leram __?

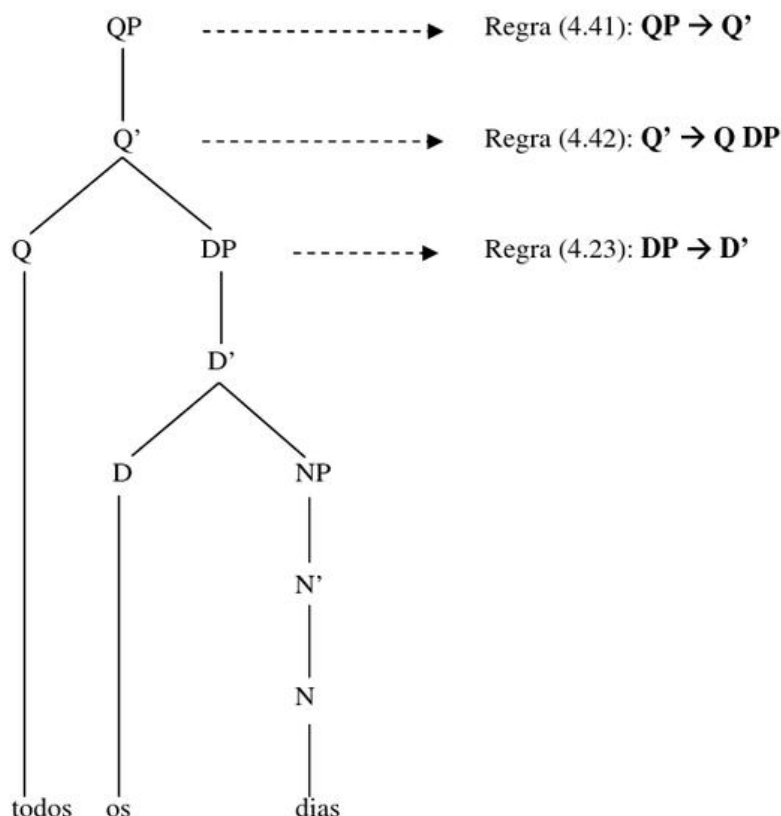
Ora, o fato que não foi percebido por Costa é que aqui não temos um DP, mas um QP[72]. E isso, por algum motivo, torna o julgamento de gramaticalidade mais confuso. A interpretação que sugerimos é que referencialmente um QP parece estar entre um DP definido e um DP indefinido, o que nos leva à escala de graus de gramaticalidade apontados por Costa:

(4.105) Quais livros [o João] leu __?
(4.106) ? Quais livros [vários/todos os alunos] leram __?
(4.107) *Quais livros [um homem] leu __?

O fato que queremos apontar aqui é que uma análise levando em consideração o QP poderia esclarecer estes dois fenômenos em português: a restrição do pronome pleno e a extração de elementos-wh com sujeitos pré-verbais.

Agora, passemos à aplicação de nossas regras (4.41) a (4.44), relativas à estrutura interna do QP. Comecemos com as seguintes análises:

(4.108) Ela sai [todos os dias].

[72] De fato, de acordo com Raposo (1992: 138), "em português, a classe dos Quantificadores inclui elementos como **muito(s)**, **pouco(s)**, **todo(s)**, **um**, **vários**, **bastantes**, entre outros" (grifos do autor).

Como as regras do QP não influenciam na estrutura interna do DP, a mesma regra permite tanto a análise de um sintagma como **[todos os dias]** quanto um sintagma como **[todos dias]**, este último com um determinante vazio. A análise da estrutura de um DP com o núcleo foneticamente vazio é a seguinte: $_{QP}[\ _{Q'}[\ _{Q}[\text{todos}]\ _{DP}[\ _{D'}[\ _{D}[\varnothing]\ _{NP}[...]]]]]]$. Um dos muitos problemas que iremos enfrentar com os quantificadores reside justamente aí, no fato de que alguns quantificadores exigem DPs com o **D vazio** (como *alguns*); outros com o **D preenchido** (como *ambos*); enquanto outros, como vimos com *todos*, são **indiferentes** em relação à presença explícita do determinante.

(4.109)

a. [Alguns alunos] não fizeram o tema.

b. *[Alguns os alunos] não fizeram o tema.

(4.110)

a. [Ambos os linguistas] chegaram à mesma conclusão.

b. ? [Ambos linguistas] chegaram à mesma conclusão[73].

---

[73] Essa sentença é perfeitamente gramatical para muitos falantes.

(4.111)

a. [Todos os crocodilos] são répteis.

b. [Todos linguistas] já leram um texto do Chomsky.

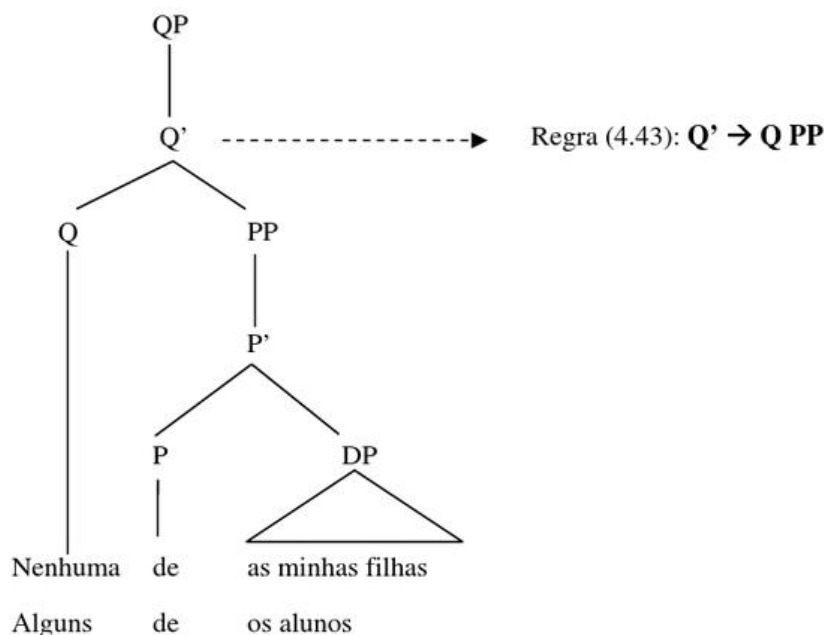
Como vimos, analisamos esse tipo de sintagma como sendo um QP (que têm um núcleo Q e que exigem um complemento DP); e não como um DP que apresenta um QP em posição de especificador, [Spec, DP], como nos sugeriu Alencar (2008) e como foi proposto por Jackendoff (1977), por exemplo. Uma vez que estabelecemos que o DP seja o complemento de Q, talvez tenhamos aí mesmo apontado para uma solução do problema da seleção categorial dos QPs. Os núcleos dos constituintes podem **s-selecionar** e **c-selecionar** seus complementos. E os núcleos funcionais apenas **c-selecionam** seus complementos[74]. Logo, podemos classificar os núcleos Q em três diferentes tipos, de acordo com sua exigência de complementação categorial: (i) há núcleos Q que exigem DPs com o D foneticamente preenchido (como *ambos*); (ii) há núcleos Q que exigem DPs com o D foneticamente vazio (como *alguns*[75]); e, finalmente, (iii) há núcleos Q que têm a informação de c-seleção subespecificada, ou seja, essa informação não é importante, uma vez que aceitam os dois tipos de DP (como *todos*).

Passemos à regra (4.43), que apresenta como complemento do quantificador um sintagma preposicional. Vejamos as análises abaixo.

(4.112) [Nenhuma das minhas filhas] vai sair hoje.

(4.113) [Alguns dos alunos] gostaram das aulas.

[74] Cf. Culicover (1997), Mioto et al. (2004), entre outros.

[75] Na verdade, o quantificador *alguns* também aceita complementos PPs, como veremos em seguida.

Os quantificadores *nenhum* e *alguns* c-selecionam PPs e não DPs. Isso fica claro com a representação arbórea que mostramos acima. Os sintagmas **[nenhuma as minhas filhas]** e **[alguns os alunos]** são agramaticais em português. Já vimos que *alguns* pode c-selecionar também um DP com o núcleo não pronunciado, como em **[alguns alunos]**. O quantificador *nenhum* parece ser um pouco mais complexo, visto que ele c-seleciona um PP (como em (4.114)) ou um DP com o traço **[+sing]** (exemplo (4.115)):

(4.114) [Nenhum dos alunos] fez o tema.
(4.115) [Nenhum aluno] fez o tema.
(4.116) *[Nenhum/nenhuns alunos] fez/fizeram o tema.

Nossas regras gramaticais, assim como estão, permitem a hipergeração dessas estruturas. No entanto, acreditamos que o problema não esteja nas regras em si, tal como as formulamos. Mais uma vez, um trabalho de descrição do léxico em termos de traços de subcategorização dos núcleos dos sintagmas pode limitar o poder de descrição das regras sintagmáticas e se beneficiar delas para a descrição dos sintagmas em PB[76].

[76] Um outro problema, de que ainda não falamos, aparece com o quantificador *todos*. Ele pode aparecer em duas posições: antes do DP ou depois do DP, em uma posição típica de adjetivo pós-nominal. Isso em se tratando de sua posição dentro do DP. Ele também pode aparecer após o verbo auxiliar (*Os estudantes estão **todos** gostando de Sintaxe*) ou após o verbo principal (*Os estudantes estão gostando **todos** de Sintaxe*). Nossas regras, infelizmente, não dão conta dessas análises. O assunto é tratado com algum detalhe por Vicente (2006).

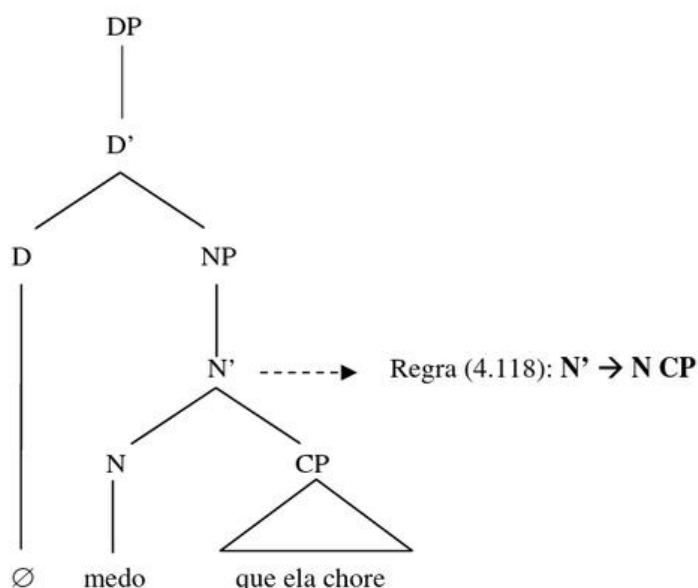
4.1.1 Sintagmas nominais com modificadores e complementos oracionais

Os sintagmas nominais também podem ter modificadores e complementos oracionais. Para dar conta desse tipo de análise, elaboramos as seguintes regras:

(4.117) N’ → N’ CP

(4.118) N’ → N CP

Os Ns transitivos podem aceitar complementos oracionais, como vemos abaixo:

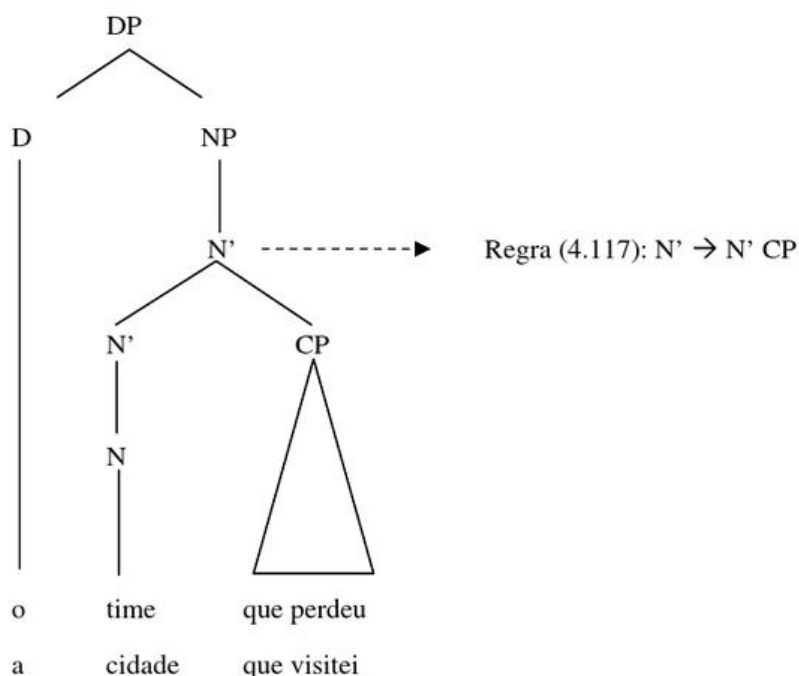
(4.119) Ele tem [medo que ela chore].

Os complementos oracionais podem também ser precedidos por uma preposição, tal como já havíamos previsto por regras anteriores[77]:

[77] Em teoria gerativa, a ausência da preposição antes dos complementos nominais oracionais se deve ao fato de que os CPs não precisam receber Caso. Logo, não precisam de uma preposição que lhes atribua/licencie o Caso (cf. Chomsky, 1986a, Raposo, 1992, entre outros). Estudos de gramaticalização do português e do espanhol chamam o fenômeno de *dequeísmo*. De acordo com Moraes de Castilho (2006: 183), “o processo de **dequeísmo** (aparecimento da preposição **de** antes da conjunção **que**) está ligado ao processo de **queísmo**, e é encontrado somente em duas línguas românicas: espanhol e português. Ele se acha em expansão nas duas variantes americanas dessas línguas, mas parece pouco produtivo nas respectivas variedades ibéricas. Existe nessas línguas desde os séculos XII-XIII, tendo ganhado relevância a partir do século XVII” (grifos da autora).

(4.120) Ele tem [medo **de** que ela chore].

Nossas regras gramaticais agora já dão conta dos dois tipos de estruturas, uma vez que já havíamos previsto um PP como complemento de um N (veja a seção anterior). Além de os CPs poderem desempenhar o papel de complementos no NP, eles também podem desempenhar o papel de adjuntos. Vejamos uma análise com nossa regra (4.117) e, em seguida, quatro argumentos para analisarmos CPs como adjuntos do NP.

(4.121) [O time que perdeu] está fora.

(4.122) [A cidade que visitei] estava linda.

Os argumentos para analisar CPs como os que vimos acima como adjuntos são basicamente os seguintes (os exemplos para cada ponto seguem logo abaixo):

(i) o NP pode apresentar mais de um CP à direita do núcleo, ou seja, há uma aplicação recursiva de CPs modificando o NP, uma característica típica de adjuntos;

(ii) o CP pode coocorrer com outros adjuntos, como APs e PPs;

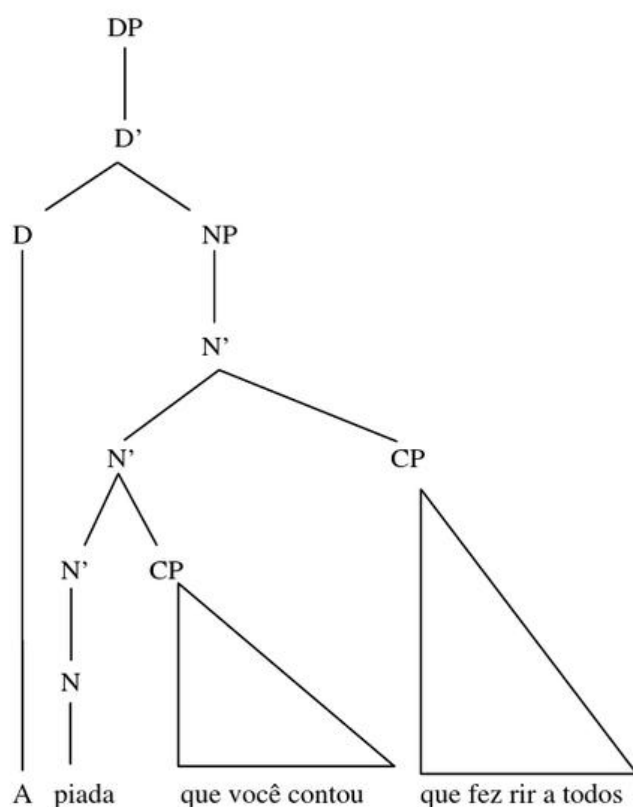
(iii) o CP pode coocorrer com Ns transitivos com complementos já expressos, ou

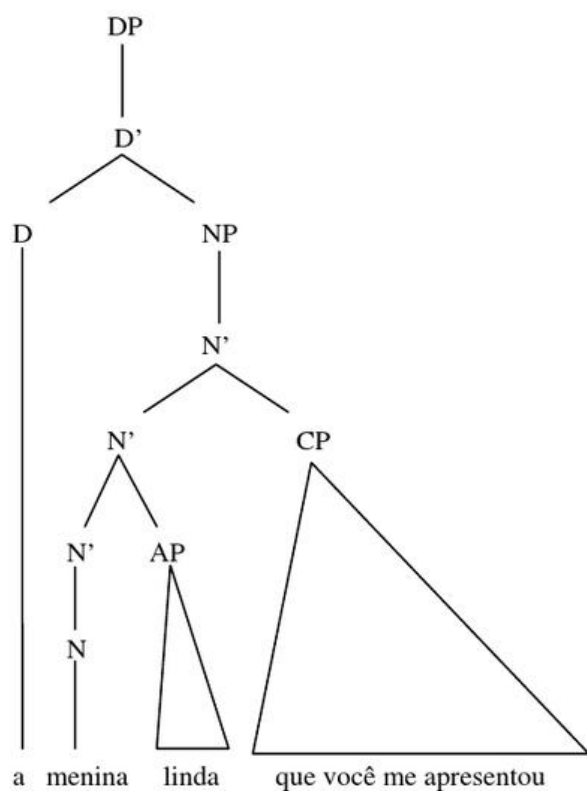
seja, com a posição de complemento já "saturada";

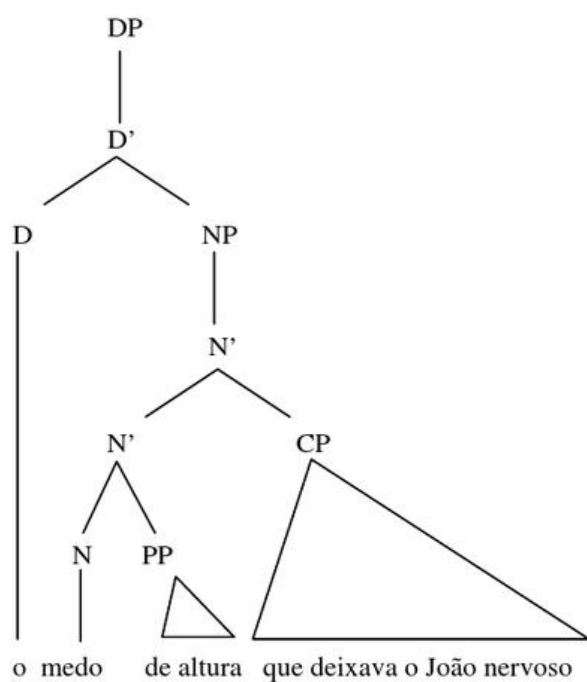
(iv) o CP adjunto pode aparecer modificando um N intransitivo, ou seja, um N que não necessita ou aceita complementos.

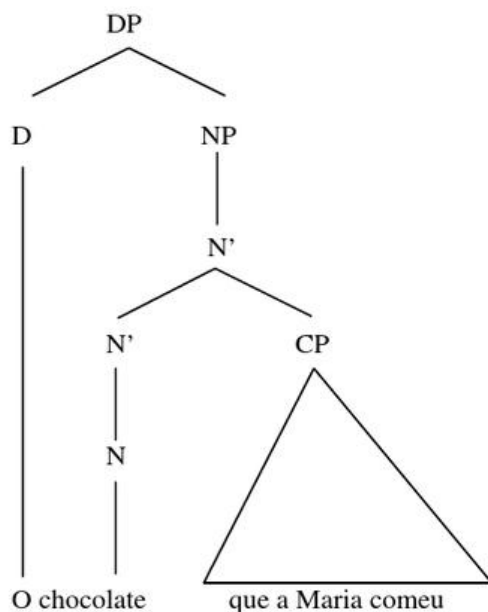
Vejamos uma análise ilustrando cada um desses pontos:

(i) reiteração de NPs adjuntos:

(4.123) [A piada que você contou que fez rir a todos] é muito engraçada.

(ii) CP e AP como adjuntos de NP:

(4.124) [A menina linda que você me apresentou]

(iii) CP adjunto com N transitivo e PP complemento:

(4.125) [O medo de altura que deixava o João nervoso]

(iv) CP adjunto modificando N intransitivo:

(4.126) [O chocolate que a Maria comeu]

Passemos a alguns problemas que encontramos na descrição da estrutura sintática dos sintagmas nominais.

4.1.2 Alguns problemas na descrição dos sintagmas nominais

Certamente nossas regras não serão capazes de descrever **todas** as estruturas do sintagma nominal em português. Teremos algum mérito se as regras gramaticais que apresentamos conseguirem prever muitas das estruturas bem formadas do sintagma nominal e conseguirem dar conta de uma ampla variedade de NPs em PB. Já vimos apontando algumas limitações de nossas regras gramaticais para o NP ao longo do capítulo e já discutimos alguns problemas nas seções anteriores. Nesta seção, discutiremos alguns outros problemas que não conseguimos resolver com as regras gramaticais que apresentamos até agora.

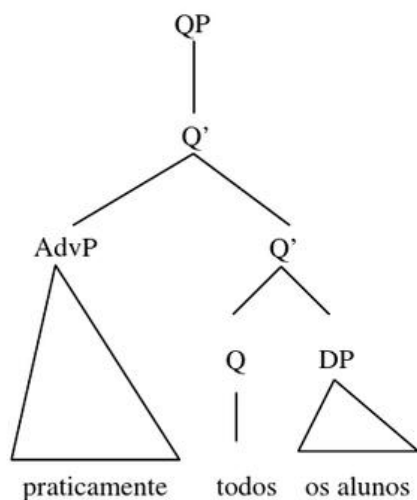
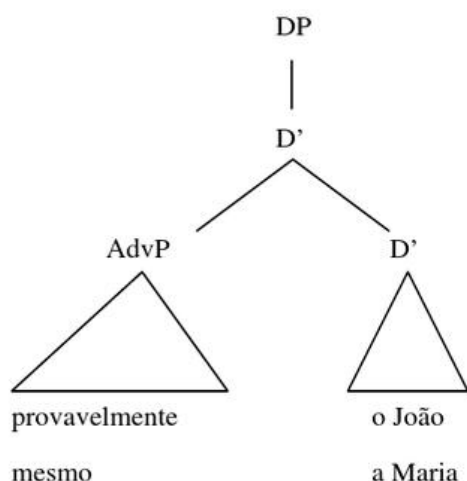
4.1.2.1 Advérbio como modificador nominal

Repare nas seguintes sentenças:

(4.127) [Praticamente todos os alunos] fizeram o tema.

(4.128) [Provavelmente o João] doou os jornais para a biblioteca. (não a Maria)[78]

(4.129) [Mesmo a Maria] sabe que tem poucas chances de ganhar o campeonato municipal[79].

Seguindo nossa intuição – tanto como falantes nativos, bem como linguistas –, parece-nos que, nessas frases, o advérbio está modificando o agrupamento nominal (o QP em (4.127) e o DP em (4.128) e (4.129)). Poderíamos analisar essas estruturas da seguinte maneira:

[78] Exemplo de Mioto et al. (2004: 17).
[79] Exemplo de Klein (2003: 139).

Para efetuarmos essas análises, podemos imaginar regras como as seguintes:

Q' → AdvP Q' ou **QP → AdVP QP**
D' → AdvP D' ou **DP → AdvP DP**

Contudo, acreditamos que essas regras sejam muito ousadas, basicamente por dois motivos. O primeiro porque elas são recursivas, o que permite que analisem sequências em que exista mais de um AdvP modificando um QP ou DP, resultando em sequências claramente agramaticais como **provavelmente praticamente o João*[80]. O segundo motivo é porque estaríamos afirmando aqui que um AdvP **pode modificar um agrupamento nominal**, contrariamente ao que postula a tradição gramatical (e também contrariamente a alguns estudos gramaticais recentes, que argumentam que a modificação adverbial se dá em um nó **FocusP** ou algo do gênero).

Tradicionalmente, aceita-se que o advérbio (ou o adjunto adverbial) possa modificar apenas o verbo, o adjetivo ou algum outro elemento adverbial. Bechara (1960: 62), por exemplo, afirma que "o adjunto adverbial é expresso por advérbio ou locução adverbial que pode referir-se não só ao verbo, mas ainda ao adjetivo e a outro advérbio"; Rocha Lima (1984: 153) diz que "advérbios são palavras modificadoras do verbo. (...) Alguns advérbios, chamados de *intensidade*, podem também prender-se a adjetivos, ou a outros advérbios, para indicar-lhes o grau". (grifos do autor).

A ideia de que um AdvP possa modificar um DP, no entanto, faz-nos sentido.

[80] No entanto, repare nesta sequência, em que os advérbios parecem estar adjungidos ao QP: [praticamente [quase [todos os alunos]]] estudaram.

Mioto et al. (2004: 17) atentam para o fato de que um AdvP pode modificar um DP. Jackendoff (1972) mostra como os advérbios *even*, *only* e *just* podem modificar um DP para torná-lo foco da sentença (esse mesmo assunto é tratado mais tarde por Klein, 2003, para o PB). Perini (2006) também admite que um advérbio (ele usa o exemplo do advérbio *só*) pode atuar como modificador de substantivo. E Lemle (1984: 168) aponta para o mesmo fenômeno (para ela, a modificação adverbial talvez seja uma propriedade da palavra nominal e não de uma classe de advérbios):

> até mesmo a classe dos nomes, predominantemente modificada por adjetivo, tem um subconjunto, semanticamente caracterizado como referente a eventos, que admite modificação por advérbios (...) Os exemplos são [os seguintes]:
> 291
> a. a nossa revolta hoje
> b. o desembarque aqui
> c. a votação agora
> d. a chegada lá
> e. uma declaração assim

Alguns estudos de cunho funcionalista em português também concordam que o advérbio pode modificar elementos nominais, como Moura Neves (2000) e Pezatti & Nogueira (2005), por exemplo.

Costa & Costa (2001), apesar de seguirem a tradição gramatical e declararem que "um advérbio nunca modifica um nome, sendo essa função desempenhada por adjectivos (...) **Os advérbios podem modificar qualquer categoria não-nominal**" (p. 76, grifos dos autores), acabam mostrando advérbios que, a nosso entender, podem modificar um agrupamento nominal. Os exemplos abaixo são extraídos-dos ou inspirados-nos exemplos de Costa & Costa (2001: 32-33), com nossa numeração. A classificação taxonômica dos advérbios é de Costa & Costa (2001):

. Advérbios de afirmação:

(4.130) O Pedro leu [certamente o livro], não a revista.

. Advérbios de dúvida:

(4.131) [Provavelmente o Pedro] leu o livro, já que o João não lê.

. Advérbios de negação:

(4.132) O Pedro leu [não o livro], mas a revista.

. Advérbios de ordem:

(4.133) O Pedro leu [primeiramente o livro] e depois, a revista.

. Advérbios de exclusão:

(4.134) O Pedro leu tudo, [exceto o livro].

. Advérbios de inclusão:

(4.135) O Pedro leu [inclusive o livro].

. Advérbios de designação:

(4.136) [Eis o livro].

. Advérbios orientados para o falante:

(4.137) [Felizmente o Pedro] leu o livro, e não a Maria, que é uma chata.

. Advérbios de foco:

(4.138) O Pedro leu [só o livro]; ele nem se interessou pela revista e pelo jornal.

. Advérbios de realce:

(4.139) [Até o Pedro] leu o livro.

Nossa intuição aponta fortemente para o fato de que, em todos esses exemplos, os advérbios estão modificando um substantivo. Ainda assim, apesar de todos esses exemplos terem sido baseados nos exemplos e na classificação de Costa & Costa (2001), esses autores não estipulam nem endossam qualquer regra que preveja a modificação nominal por um agrupamento adverbial[81]. Também por isso, essa intuição que temos de que os advérbios podem modificar um DP não será expressa pelas regras que formulamos aqui. Tal regra acabaria prevendo muitas estruturas agramaticais (ou, pelo menos, inaceitáveis):

(4.140) *[Sinceramente o João] pediu desculpas à Maria.

[81] João Paulo Cyrino, em comunicação pessoal, concorda com a análise de Costa & Costa, justificando detalhadamente qual constituinte cada um desses advérbios modifica (nunca o DP ou NP). Magdiel M. Aragão Neto, também em comunicação pessoal, também discorda de uma interpretação de modificação adverbial dos agrupamentos nominais nesses exemplos. Isso reforça a ideia de que a modificação nominal por um elemento adverbial não é um consenso entre os sintaticistas.

(4.141) *[Ontem o Grêmio] jogou bem.

Quer dizer, não entendemos que a modificação adverbial em agrupamentos nominais seja uma regra produtiva do português. Parece-nos que apenas algumas classes de advérbios possam atuar sobre um DP ou sobre um QP (ou, de maneira oposta, como sugere Lemle (1984), apenas algumas classes de substantivos aceitam modificação adverbial). Isso é mais um indício que sugere que a classe dos elementos que chamamos de "advérbio" é extremamente heterogênea. É muito comum encontrarmos frases como a seguinte, de Perini, em trabalhos sobre o estudo dos advérbios: "Chamar a todos [os advérbios tal como se apresentam na NGB] de 'advérbios' é ignorar um mundo de fenômenos gramaticais importantes" (2006: 163-4). Por isso, aceitaremos por ora essa limitação em nossas regras gramaticais, mesmo que, intuitivamente, acreditemos que alguns advérbios possam modificar um agrupamento nominal.

4.1.2.2 Elementos pesados no sintagma nominal

Outro problema que encontramos no trabalho de descrição do sintagma nominal em português foi com a ordem de elementos pesados dentro do NP. Vejamos as seguintes sentenças.

(4.142)

a. $_{DP}$[O homem $_{AP}$[orgulhoso]] descansa em paz.

b. $_{DP}$[O $_{AP}$[orgulhoso] homem] descansa em paz.

(4.143)

a. $_{DP}$[O homem $_{AP}$[orgulhoso de seus feitos]] descansa em paz.

b. */?? $_{DP}$[O $_{AP}$[orgulhoso de seus feitos] homem] descansa em paz.

Um sintagma adjetival pode ser modificador de um sintagma nominal, como já explicitamos acima, com as regras (4.30) e (4.31) e como vemos em (4.142). No entanto, se o AP for "pesado", ele não pode modificar o NP em posição pré-nominal, apenas em posição pós-nominal, como vemos em (4.143).

Não temos como prever isso com nossas regras gramaticais. Provavelmente, nesses casos esteja atuando algum princípio de processamento psicolinguístico, como o EIC (do inglês *Early Immediate Constituents*). O princípio diz que nosso mecanismo de *parsing* mental prefere organizar os elementos da melhor maneira possível para o processamento da informação, o que implica "fechar" um constituinte o mais próximo possível do núcleo. Sabemos que, em línguas SVO, os elementos mais pesados costumam vir **depois** dos elementos leves[82].

O mesmo efeito de processamento acontece nas sentenças abaixo:

(4.144)

a. $_{DP}$[A menina [linda] [que você me apresentou]] me ligou.

b. */?? $_{DP}$[A menina [que você me apresentou] [linda]] me ligou.

(4.145)

a. $_{DP}$[O jogador [do Grêmio] [que se machucou]] chegou.

b. */?? $_{DP}$[O jogador [que se machucou] [do Grêmio]] chegou.

Mais uma vez, o elemento "pesado" só é bem acomodado na parte final do constituinte. O problema já é conhecido na literatura, mas não temos como lidar com ele utilizando apenas as regras categoriais de reescrita livres de contexto que apresentamos.

Passemos ao estudo do sintagma adjetival.

4.2 O SINTAGMA ADJETIVAL

Assim como fizemos para elaborar as regras do NP, ao começar nosso estudo do sintagma adjetival, partimos de análises anteriores[83]. Elaboramos as seguintes regras de reescrita que, acreditamos, são capazes de prever a estrutura dos APs bem formados em português.

---

[82] Cf. Hawkins (1994, 1997), Hoffman (1999) e Newmeyer (s.d). O curioso é que essa mesma intuição também pode ser encontrada em alguns estudos de ordenamento sintático de gramáticos tradicionais do português. Pereira (1958[1918]), por exemplo, recomenda que complementos relativos longos devem aparecer sempre em posição pós-nuclear, enquanto complementos curtos podem aparecer em posição pré-nuclear.

[83] Nossas análises do AP baseiam-se em grande parte em Lemle (1984), Perini (2000) e Othero (2006).

(4.146) AP → A'

(4.147) A' → A

(4.148) A' → AdvP A'

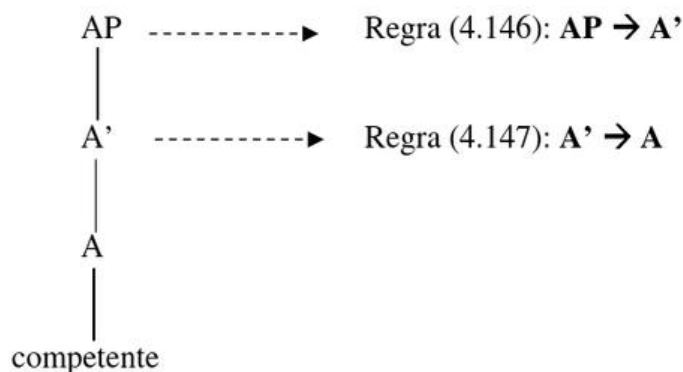
(4.149) A' → A' AdvP

(4.150) A' → A' PP

(4.151) A' → A PP

(4.152) A' → A CP

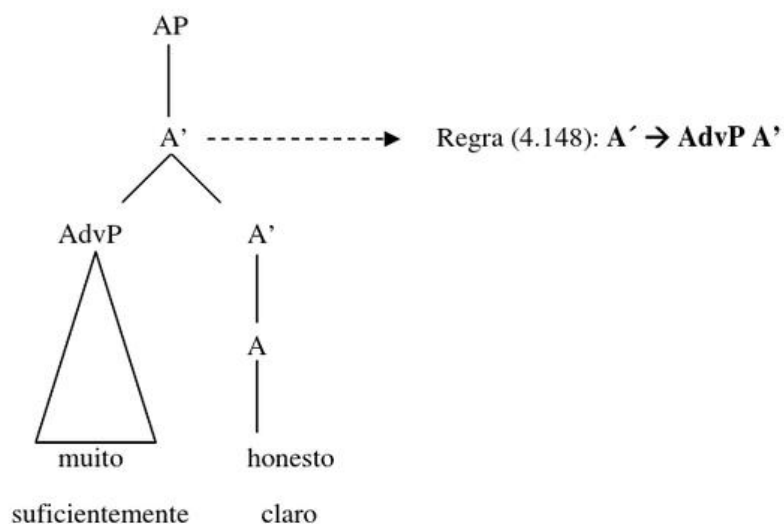
Vejamos como analisar os seguintes sintagmas adjetivais:

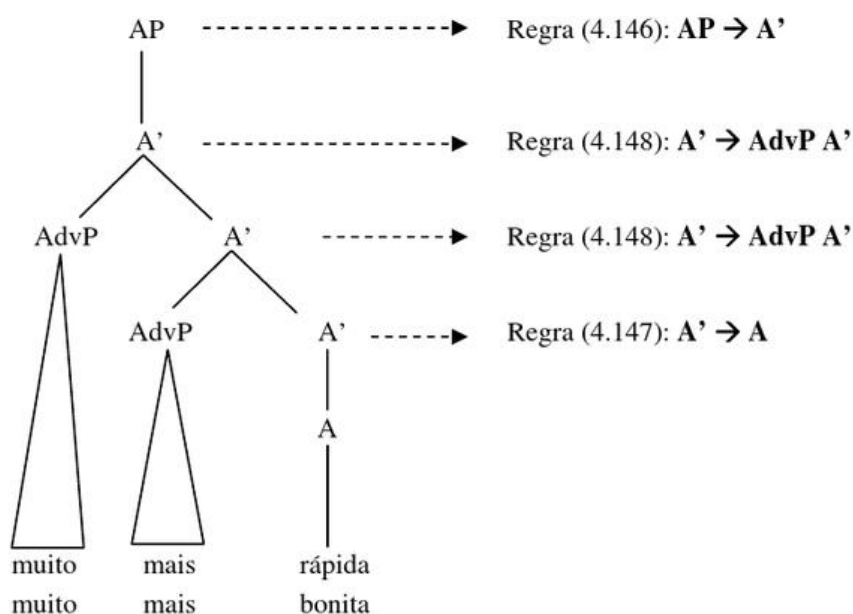
(4.153) João é um professor [competente].

Essas duas regras descrevem um AP constituído apenas por um núcleo adjetival. As regras seguintes descrevem um AP com modificação adverbial. Estamos encarando o AdvP a seguir como o **adjunto** do AP e não como seu especificador. Assumiremos (seguindo Fukui & Speas, 1986; Raposo, 1992; Mioto et al., 2004, entre outros) que a posição do especificador do sintagma adjetival fica vazia e que os modificadores adverbiais do AP estão em posição de adjunção, assim como também fizemos com as análises do DP. Os dois argumentos tradicionais em favor dessa hipótese são os seguintes: (i) os adjuntos, ao contrário dos especificadores, podem ser reiterados dentro do sintagma (ver exemplos abaixo); e (ii) em português, a posição prototípica do **Spec** é à esquerda do núcleo $\mathbf{X^0}$; no entanto, os modificadores adverbiais podem aparecer tanto à esquerda como à direita de $\mathbf{X^0}$.

Vejamos os sintagmas abaixo:

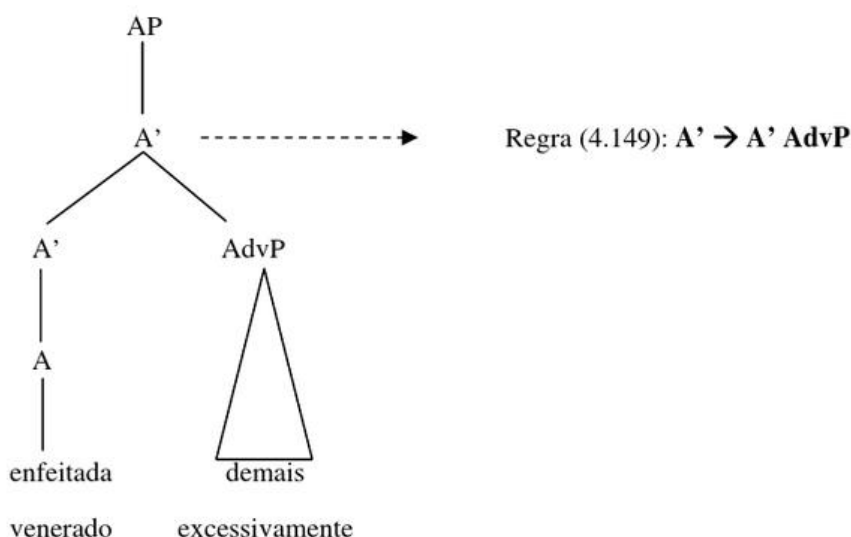
(4.154) João é [muito honesto].

(4.155) Isso está [suficientemente claro].

Aqui, vemos um AP com modificação adverbial à esquerda do núcleo.

A mesma regra pode ser aplicada mais de uma vez, para analisar sintagmas com mais de um modificador adverbial.

(4.156) Essa máquina é [muito mais rápida].
(4.157) Maria é [muito mais bonita].

O modificador adverbial também pode aparecer à direita do núcleo adjetival. Veja os sintagmas abaixo:

(4.158) Maria está [enfeitada demais].

(4.159) João é [venerado excessivamente].

Essas duas regras, no entanto, não estão livres de problemas. Elas são regras que predizem e analisam estruturas bem formadas dos sintagmas adjetivais em português e podem ser entendidas como regras que refletem um processo de formação produtivo na língua, qual seja: os sintagmas adjetivais podem ser modificados por elementos adverbiais. No entanto, encontramos dois problemas nessas regras.

O primeiro problema é o fato de elas permitirem que um sintagma adjetival seja modificado por mais de um modificador adverbial **à direita** do núcleo. Interessantemente, os modificadores adverbiais podem ser reiterados à esquerda, mas não à direita do núcleo. Repare na agramaticalidade dos seguintes sintagmas:

(4.160)

a. Maria é $_{AP}$[bonita].

b. Maria é $_{AP}$[muito bonita].

c. Maria é $_{AP}$[realmente muito bonita].

(4.161)

a. Maria é $_{AP}$[bonita demais].

b. *Maria é $_{AP}$[bonita demais excessivamente].

c. *Maria é $_{AP}$[bonita excessivamente demais].

Não sabemos como lidar com esse problema. Podemos supor como uma hipótese inicial que os modificadores adverbiais tenham um comportamento idiossincrático em relação à sua posição no AP. Se estivermos certos, podemos pensar que os itens lexicais devem receber traços sobre sua posição como modificadores adverbiais, mais ou menos como postulou Perini (1989, 2000) para tratar da ordem dos elementos modificadores dentro dos sintagmas (ver discussão mais abaixo). Esses traços lexicais poderiam "barrar" a reiteração de elementos modificadores à direita do núcleo.

Outro problema que está relacionado a esse é o seguinte: as regras que propusemos até agora permitem a formação de sintagmas adjetivais com dupla modificação adverbial. Ou seja, elas preveem a boa formação de um sintagma adjetival que seja modificado por um AdvP, tanto à esquerda como à direita do núcleo adjetival, o que não acontece em PB. Veja os sintagmas abaixo:

(4.162) *$_{AP}$[$_{A'}$[$_{AdvP}$[muito] [$_{A'}$[$_{A'}$[bonita] $_{AdvP}$[demais]]]]

(4.163) *$_{AP}$[$_{A'}$[$_{AdvP}$[excessivamente] [$_{A'}$[$_{A'}$[claro] $_{AdvP}$[demais]]]]

Tal estrutura é agramatical em português[84]. Lemle (1984), ao postular sua regra de descrição dos sintagmas adjetivais, não prestou atenção ao problema e apresentou a seguinte regra:

SAdj → (Adv) Adj (Adv) (SPrep)* (S)*[85]

A regra de Lemle (1984), assim como as nossas, prevê que um sintagma adjetival possa ser modificado por dois elementos adverbiais simultaneamente, um à esquerda do núcleo e outro à direita do núcleo. Os fatos da língua, no entanto, nos mostram que essa

[84] Cf., no entanto, a seguinte frase: "Maria é [realmente bonita demais]", que apresenta dois modificadores adverbiais ligados ao agrupamento adjetival. Agradecemos a Mathias Schaff, em comunicação pessoal, por este exemplo.

[85] Esta é a regra de número 9 (p. 153) de Lemle. **SAdj** é o sintagma adjetival; **SPrep** é o sintagma preposicional. Os parênteses indicam a opcionalidade do elemento; o asterisco indica a recursividade dos elementos.

não é uma estrutura bem formada em PB[86]. Infelizmente, nossas regras categoriais de reescrita são baseadas em um formalismo de uma gramática livre de contexto, como vimos no capítulo 2. Por isso, não conseguiremos elaborar uma regra que traduza nossa intuição de que, se houver um elemento modificador adverbial à esquerda do núcleo, não poderá haver um elemento de mesma natureza à direita, e vice-versa. Essa é mais uma limitação que teremos de aceitar em nossa gramática tal como a apresentamos até agora. Concordamos com Perini (2000: 113) quando diz que "a estrutura interna do SAdj encerra alguns mistérios, que estão ainda à espera de estudos aprofundados". A regra que Perini apresenta para a descrição do sintagma adjetival traz essa mesma limitação (veja adiante).

Em relação à **ordem** dos modificadores adverbiais, podemos postular a existência de ao menos três diferentes classes[87]: (i) alguns modificadores adverbiais podem aparecer apenas em **posição pré-adjetival** (à esquerda do núcleo); (ii) alguns somente em **posição pós-adjetival** (à direita do núcleo); (iii) enquanto outros podem aparecer nas **duas posições** (mas não simultaneamente, como vimos acima). Veja os seguintes exemplos:

(4.164) Maria estava [**muito** bonita] / *[bonita **muito**].
(4.165) Maria estava [bonita **demais**] / *[**demais** bonita].
(4.166) Maria estava [bonita **exageradamente**] / [**exageradamente** bonita][88].

Perini (1989, 2000) traz um estudo detalhado a respeito da ordem dos elementos modificadores do sintagma adjetival. Para ele, a estrutura do sintagma adjetival pode ser descrita pela seguinte regra[89]:

[Int-4 Int-3 Int-2 NSA Int-1 CSA]

---

[86] Repare também que a regra de Lemle (1984) não prevê a recursividade do modificador adverbial dentro da estrutura do AP. Por um lado, sua regra formaliza uma intuição de maneira equivocada, uma vez que **podemos ter mais de um AdvP** modificando o AP **à esquerda** do núcleo; por outro lado, sua regra formaliza uma intuição correta, a de que é possível termos apenas **um** AdvP modificando o AP à direita do núcleo.

[87] Perini (2000) sugere uma divisão em quatro classes de modificadores adjetivais. Não concordamos com essa divisão especialmente por dois motivos: (i) Perini considera elementos que nos parecem ser modificadores verbais e adverbiais, e não adjetivais, como *[**realmente sempre** muito satisfeito com a vida]* (p. 114); (ii) a classificação de Perini perde em generalidade, pois tenta estipular nas regras sintagmáticas restrições de posição para elementos de mesma natureza – entendemos que essas restrições poderiam aparecer como **traços lexicais** – ver discussão a seguir.

[88] Ignoramos se a modificação pré-nuclear ou pós-nuclear possa trazer modificações de natureza semântica.

[89] Cf. Perini (2000: 115). A regra de Perini, assim como a de Lemle (1984) e a nossa, prevê erroneamente que um AP possa ter simultaneamente dois modificadores adverbiais, um à direita e outro à esquerda do núcleo.

Int é o intensificador, sempre um advérbio. O número corresponde à posição que ele pode aparecer em relação ao núcleo do AP (para Perini, **NSA**). Acreditamos que tal regra seja por demais específica e perca em generalização. Perini sobrecarrega a regra gramatical por não contar com o trabalho "intralexical", com traços lexicais. A informação sobre a posição do modificador adverbial (que pode aparecer, como vimos, **antes** do núcleo e **depois** do núcleo) deve ser uma informação lexical que não se reflete nas regras de descrição gramatical. Aparentemente há muita idiossincrasia com relação à posição de cada modificador[90]. Perini assume que

> a distribuição dos itens léxicos que podem desempenhar as quatro funções denominadas 'intensificadores' é relativamente complexa. Existem, no mínimo, quatro classes de itens: os que só podem ser Int-1 (demais); os que só podem ser Int-2 (muito); os que podem ser Int-3 ou Int-1 (sempre); e os que podem ser Int-4 ou Int-1 (realmente). Os itens destas duas últimas classes parecem ser relativamente numerosos; já demais e muito são, aparentemente, representantes únicos de suas classes (...).

Entendemos que tal idiossincrasia não deva ser refletida nas regras gramaticais de boa formação dos sintagmas, pois isso acaba onerando as regras sintagmáticas da gramática e perdendo em generalidade (ao invés de propor somente uma classe "modificadores" ou "sintagmas adverbiais", como fizemos, Perini propôs quatro classes distintas: Int1, Int2, Int3 e Int4). Podemos manter as regras mais simples e mais "puras" (além de menos específicas) se expressarmos as restrições posicionais dos elementos adverbiais como traços lexicais diretamente em cada item lexical[91]. Ou seja, os itens lexicais podem ter o traço de posição 1, 2, 3 ou 4.

Passemos às três últimas regras de descrição do AP. A regra (4.150) permite-nos descrever sintagmas adjetivais com modificadores preposicionais. Esses modificadores também estão em posição de adjunção.

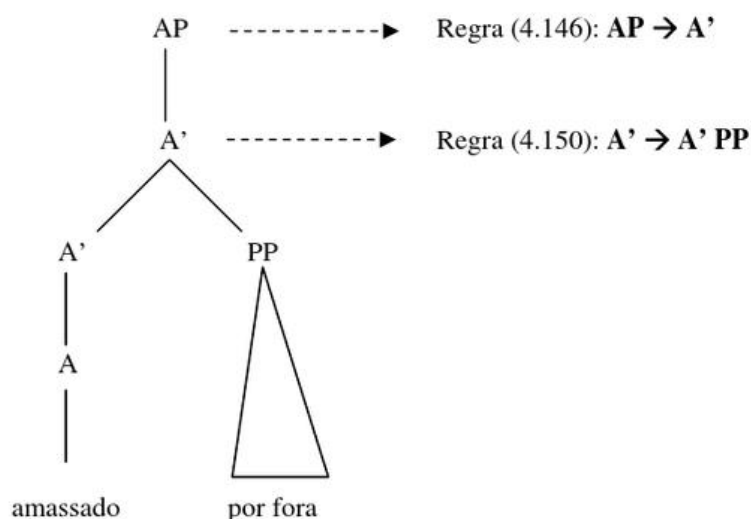
---

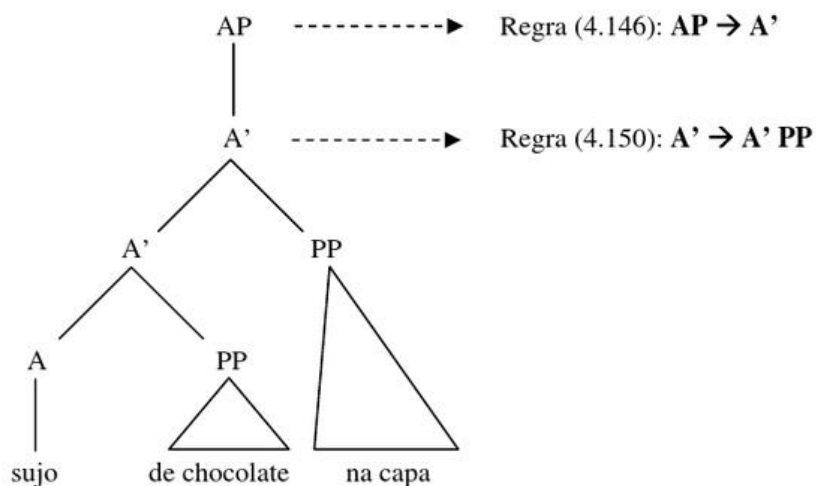
[90] Sejamos honestos: também nós não contamos com um trabalho de especificação de traços lexicais. Contudo, estamos elaborando nossas regras gramaticais prevendo um trabalho futuro de especificação lexical. Sabemos que as regras sintagmáticas que apresentamos aqui podem ser implementadas computacionalmente de maneira relativamente tranquila se contarmos com um componente lexical que apresente traços suficientemente específicos que restrinjam a aplicação de nossas macrorregras sintagmáticas. Exemplos para o português de aplicações computacionais de CFGs que contaram com um léxico ricamente trabalhado podem ser vistos em Pagani (2004), Othero (2004, 2006), David (2007) e Othero & David (no prelo).

[91] Essa discussão sobre colocar o peso da explicação nas regras gramaticais ou no léxico não é recente em gramática gerativa. Uma boa ilustração do assunto aparece no terceiro capítulo de Sag, Wasow & Bender (2003), em que os autores argumentam em favor dos traços lexicais no que diz respeito a informações de subcategorização verbal e concordância, livrando assim sua PSG livre de contexto de especificar tais informações nas regras gramaticais. Cf. também Chomsky (1970).

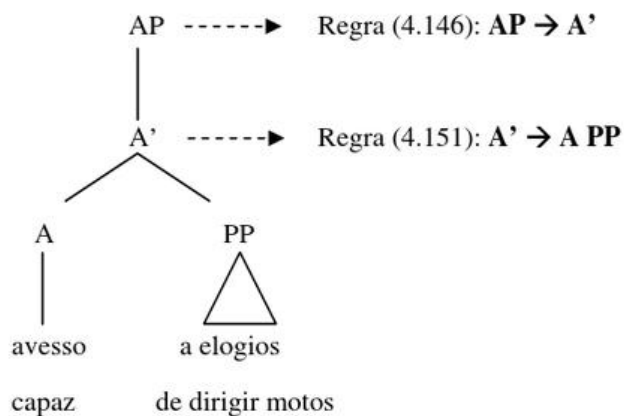
Da mesma forma que os substantivos, os adjetivos também podem apresentar complementos e adjuntos. Ainda que a distinção nem sempre se faça clara, assumimos a ideia consensual de que existem adjetivos **intransitivos**, que não necessitam complemento, e adjetivos **transitivos**, que aceitam complemento. De acordo com Lemle (1984: 153), dentro do AP, encontramos PPs que "têm a função de modificadores, nisso equivalendo a advérbios, e há os que (...) têm a função lógica de argumentos do adjetivo", ou seja, complementos do núcleo adjetival.

Comecemos com a análise dos adjetivos intransitivos modificados por PPs. Vejamos os seguintes sintagmas:

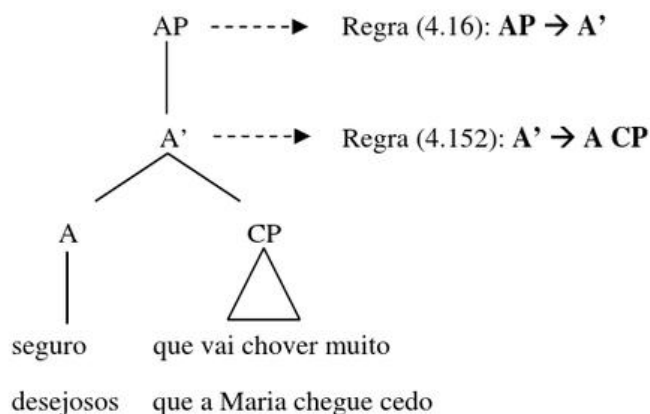
(4.167) Seu terno está [amassado por fora].

Sendo adjuntos, os PPs podem ser reiterados no AP:

(4.168) O livro está [sujo de chocolate na capa].

Vejamos a análise de adjetivos que apresentam complementos:

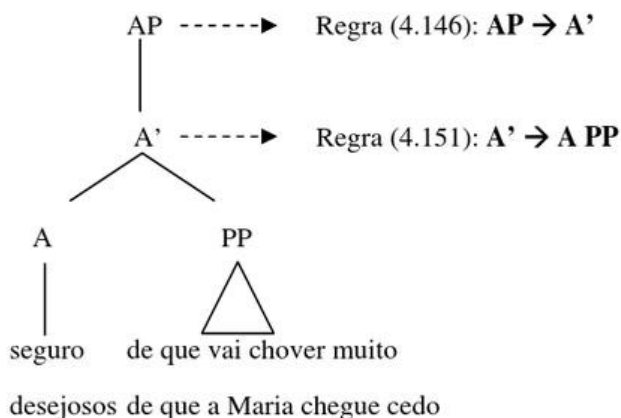
(4.169) João é [avesso a elogios].

(4.170) Maria é [capaz de dirigir motos].

Os complementos dos adjetivos podem também ser CPs. Vejamos as seguintes análises:

(4.171) Estou [seguro que vai chover muito].

(4.172) Todos estão [desejosos que a Maria chegue cedo].

O complemento oracional também pode ser precedido por preposição. Ou seja, os complementos dos adjetivos podem ser CPs ou PPs. Vejamos os exemplos seguintes:

(4.173) Estou [seguro **de** que vai chover muito].

(4.174) Todos estão [desejosos **de** que a Maria chegue cedo].

Assim como vimos com os complementos nominais oracionais na estrutura do NP, os CPs que atuam como complementos nominais no AP também podem aparecer precedidos ou não de preposição.

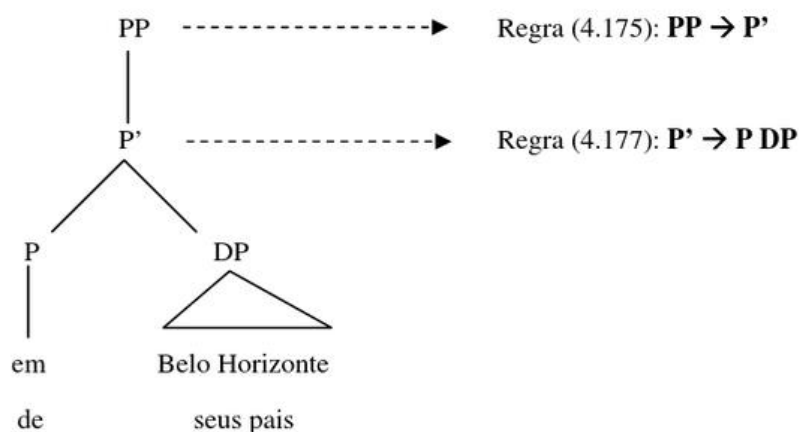
Passemos, então, à análise do sintagma preposicional.

4.3 O SINTAGMA PREPOSICIONAL

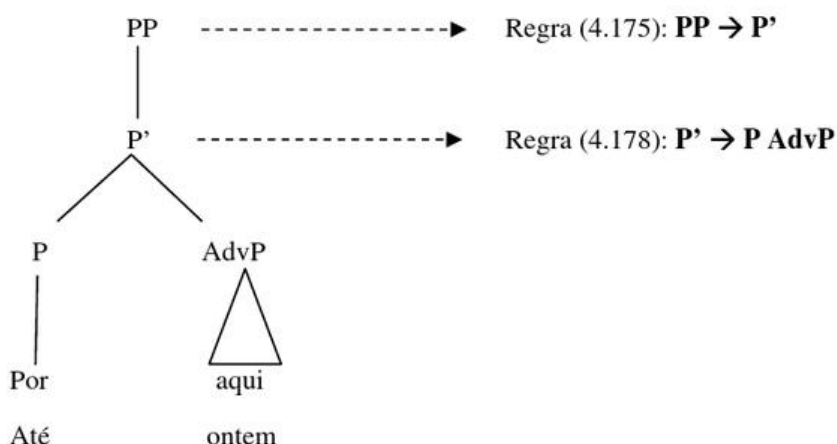
Para analisar a estrutura do sintagma preposicional, partimos das regras gramaticais que já havíamos estudado anteriormente. Com base nelas, propusemos as seguintes regras:

(4.175) PP → P'
(4.176) P' → AdvP P'
(4.177) P' → P DP
(4.178) P' → P AdvP
(4.179) P' → P CP
(4.180) P' → P PP
(4.181) P' → P

Vejamos algumas análises que argumentam em favor da elaboração dessas regras. Comecemos com a análise dos seguintes PPs:

(4.182) João nasceu [em Belo Horizonte].
(4.183) João ainda mora na casa [de seus pais].

Estamos encarando o DP como complemento da preposição. Da mesma forma, *mutatis mutandis*, com a regra (4.178), dizemos que uma preposição pode apresentar como complemento um sintagma adverbial. Vejamos alguns exemplos:

(4.184) João sempre passa [por aqui].

(4.185) [Até ontem], eu tinha escrito 136 páginas.

Nesses exemplos, os sintagmas adverbiais estão funcionando como complementos da preposição e não como seus adjuntos. Vejamos por quê. Em primeiro lugar, esses AdvPs não são opcionais (compare **Ela sempre passa por ∅./ *O João vai estudar até* ∅.). Além do mais, se os sintagmas adverbiais fossem apenas modificadores do PP ou do P', eles poderiam também ser modificadores de PPs com preposições transitivas cujo complemento é um DP, já que estariam em uma posição de adjunção. No entanto, P-barras com argumentos internos já "saturados" não podem receber adjunção de AdvP:

(4.186)

a. Ela sempre passa $_{PP}$[por lá].

b. Ela sempre passa $_{PP}$[por aquela rua].

c. *Ela sempre passa $_{PP}$[por lá aquela rua].

d. *Ela sempre passa $_{PP}$[por aquela rua lá][92].

(4.187)

a. O João vai estudar $_{PP}$[até amanhã].

b. O João vai estudar $_{PP}$[até que sua mãe saia].

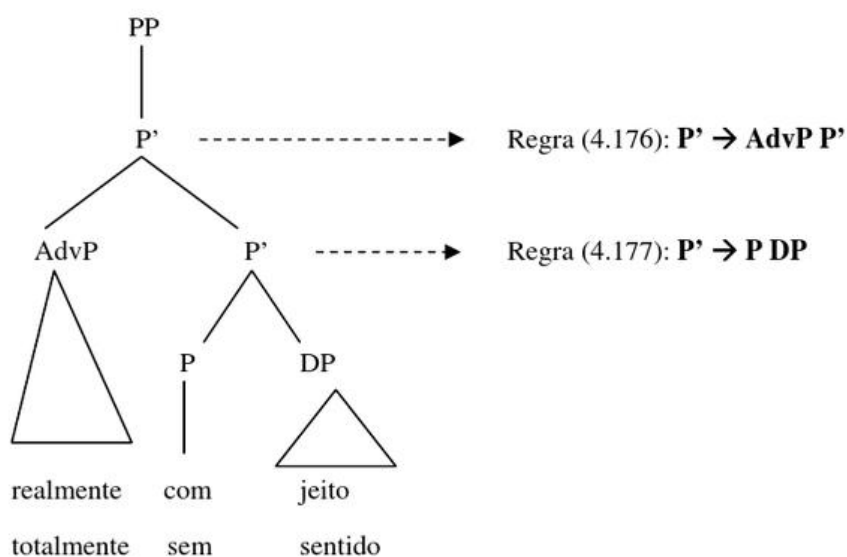
c. *O João vai estudar $_{PP}$[até amanhã que sua mãe saia].

[92] A interpretação que queremos é aquela em que tanto o DP como o AdvP completem o sentido da preposição com um "locativo", o que não parece ser possível. A frase só é gramatical se o escopo de modificação do advérbio for (curiosamente e de acordo nossa intuição) o DP, como nos mostra uma possível continuação da sentença: *Ela sempre passa por* ***aquela rua lá****, jamais anda por outra rua*. Cf. seção 4.1.2.1.

d. *O João vai estudar $_{PP}$[até que sua mãe saia amanhã][93].

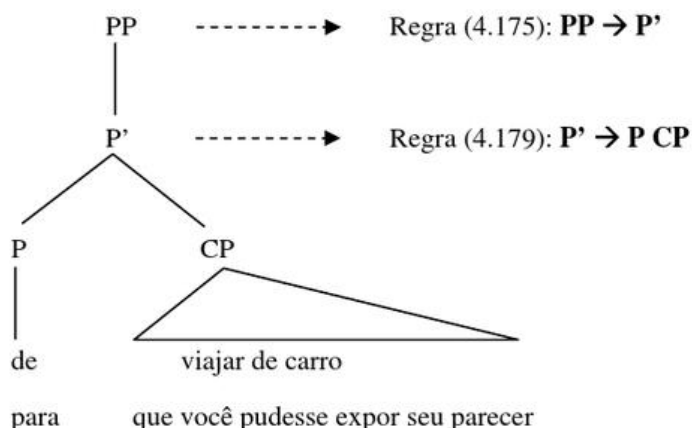
Além de figurarem em posição de complemento no PP, os sintagmas adverbiais também podem aparecer em posição de adjunção, como nos exemplos seguintes:

(4.188) Maria fez tudo [realmente com jeito].

(4.189) As atitudes de João foram [totalmente sem sentido].

Por fim, temos preposições que aceitam uma oração como complemento, um CP. Vejamos:

(4.190) João gosta [de viajar de carro].

(4.191) Todos saíram [para que você pudesse expor o seu parecer].

[93] Mais uma vez, a interpretação que queremos mostrar é aquela em que tanto o CP como o AdvP completam o sentido da preposição com um "complemento temporal", o que não parece ser possível.

Rotulamos todos esses sintagmas como **sintagmas preposicionais** independentemente da **função sintática** que o agrupamento exerce na frase. Perini (1989, 2000, 2006), por outro lado, classifica alguns de nossos PPs como sendo **sintagmas adverbiais**, por eles ocuparem uma função sintática típica de adjuntos adverbiais. Não concordamos com Perini ao classificar um PP como um AdvP. Souza e Silva & Koch (1993) adotam postura semelhante à de Perini, porém inversa: elas não reconhecem a existência do sintagma adverbial e classificam os sintagmas adverbiais como pertencentes à classe dos sintagmas preposicionais, por desempenharem "o mesmo papel: o de *modificadores circunstanciais*" (p. 19, grifos das autoras).

Nossa classificação, contrariamente à de Perini e à de Souza e Silva & Koch, segue um modelo de **rotulação categorial** dos elementos e não o de uma **rotulação funcional**. Por isso, chamaremos de sintagmas adverbiais apenas aqueles sintagmas encabeçados por um advérbio (como veremos na próxima seção) e de sintagmas preposicionais apenas aqueles sintagmas encabeçados por uma preposição, como vimos até aqui[94].

Finalmente, ao contrário do que havíamos proposto em Othero (2004, 2006), além das preposições transitivas, postulamos agora ainda uma última regra, que diz respeito às **preposições intransitivas**. Nossa proposta segue de perto os apontamentos de Lemle (1984) – que, por sua vez, seguiu a proposta de Jackendoff (1973) – e de Radford (1997a).

[94] Ernst (2001), que desenvolve um extenso trabalho sobre adjunção, apresenta dois argumentos interessantes para não confundir **AdvPs** com **PPs** em posição de adjunção. Muito resumidamente, o primeiro argumento diz respeito à ordenação entre os constituintes: enquanto os AdvPs apresentam uma ordem fixa entre eles, dependendo de sua natureza (modo, frequência, modalidade, etc.), os PPs adjuntos não apresentam restrições de ordem. O segundo argumento de Ernst diz respeito à posição em que os AdvPs e PPs adjuntos podem figurar na frase – em geral, os AdvPs podem aparecer à esquerda do núcleo, enquanto PPs, via de regra sintagmas "mais pesados", devem ocupar a posição à direita do núcleo.

Vejamos a análise dos seguintes sintagmas:

(4.192) Amanhã, o Grêmio joga [fora].

(4.193) Chomsky foi [além].

Para analisar os sintagmas destacados acima, conhecemos duas alternativas: (i) podemos aceitar o fato de que existam preposições transitivas e preposições intransitivas e classificar os sintagmas acima como PPs intransitivos; ou (ii) podemos considerar *fora* e *além* como advérbios e classificar os sintagmas acima como AdvPs.

Uma vantagem que salta aos olhos se adotarmos a primeira solução diz respeito a questões internas à teoria. Ao formularmos uma regra tal como **P' → P** (ou seja, um nível intermediário que se expanda em um núcleo lexical intransitivo, nesse caso, uma preposição), estaremos aceitando uma regra gramatical prevista pelo esquema da teoria X-barra, sem qualquer contradição ao modelo. Na verdade, a generalização se completaria agora para todos os núcleos lexicais: **XP – X' – X^0**, onde **X^0** pode ser qualquer categoria lexical (N, A, V, Adv ou P).

No entanto, uma (macro)regra como **P' → P** poderia permitir a geração de sequências claramente agramaticais, como as seguintes:

(4.194) *João gosta [de].

(4.195) *Maria saiu [com].

Mas devemos lembrar que isso não é um problema para as regras de nossa gramática. Já alertamos em algumas passagens que não iremos trabalhar com informações de traços lexicais nos itens léxicos. É evidente que nossas regras gramaticais de boa formação do sintagma preposicional supergeram estruturas. Contudo, isso pode ser remediado se pensarmos que basta postular o traço **[± transitivo]** para cada núcleo **P** para darmos conta da boa formação desses sintagmas preposicionais – o mesmo traço que deve ser implementado em todos os demais núcleos lexicais que estudamos até agora. Um traço **[± transitivo]** irá restringir a aplicação das regras sintagmáticas e barrar sequências mal formadas, como (4.194) e (4.195).

Por isso, postulamos a existência de preposições **transitivas** e **intransitivas**. As preposições de (4.192) e (4.193) são preposições intransitivas, que não precisam de

complemento, enquanto as preposições em (4.194) e (4.195) são transitivas, no sentido de que necessitam de um complemento para tornarem as sentenças gramaticais[95].

Para Lemle (1984: 161), uma grande virtude dessa proposta é que ela

> (...) tem o poder de revelar a semelhança existente entre a classe das preposições e as demais classes de palavras, pois só com esse tratamento ganhamos a generalização de que a transitividade ou intransitividade é uma propriedade de subcategorização estrita não apenas dos verbos, mas também dos nomes, adjetivos e preposições. Constatar paralelismos intercategoriais é ir ao encontro da teoria do X'.

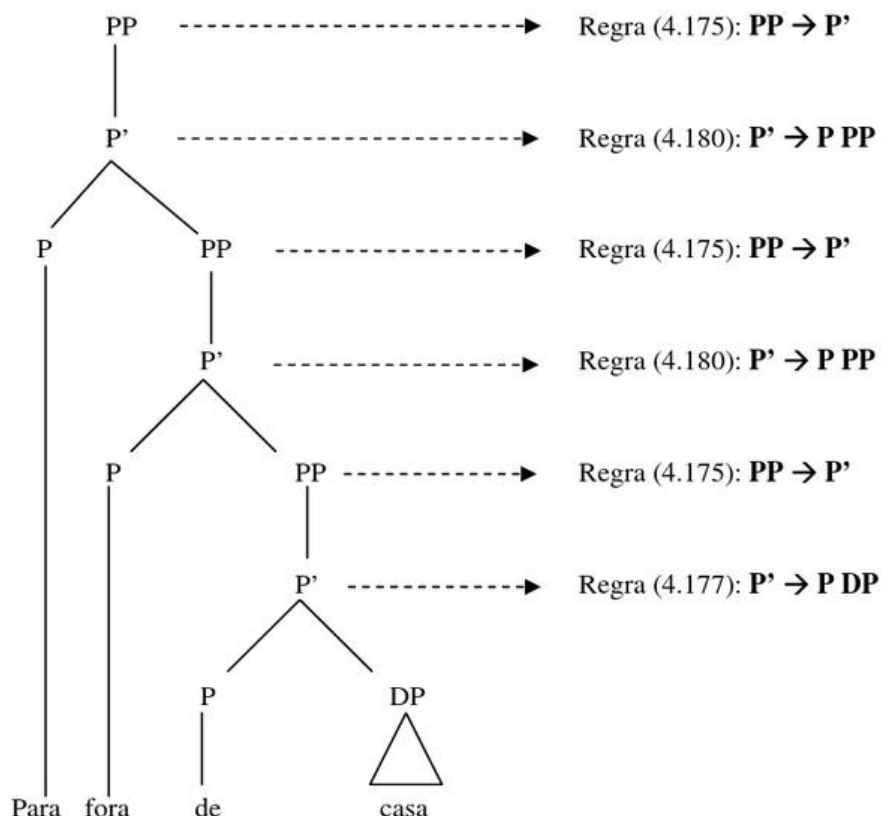
Nesse espírito, passemos à regra que analisa as **locuções preposicionais**:

(4.196) João sempre foi caridoso [para com os doentes].

(4.197) Maria colocou João [para fora de casa].

Nesses exemplos, o PP tem uma estrutura do tipo **PP[P'[P] [PP]]**. Tradicionalmente, as gramáticas normativas do português classificam esse tipo de construção como sendo uma **locução prepositiva**[96], como as chamamos acima. De acordo com nossas regras, a árvore do sintagma em (4.197) é a seguinte:

[95] Há ainda as preposições que podem ser utilizadas de maneira intransitiva se for possível retomar contextualmente seu complemento, como em *Na votação, o João foi contra* (exemplo de Sérgio Menuzzi, em comunicação pessoal) e *Ela veio com o material e eu vim sem* (Perini, 2006: 167). Esse é mais um argumento favorável à nossa decisão sobre as preposições intransitivas. Parece haver duas classes de preposição: as que **necessitam complementos** e **as que aceitam complemento**. Voltaremos a essa observação em seguida (e discutiremos com mais detalhe a transitividade dos núcleos na seção 4.5, quando tratarmos do sintagma verbal).

[96] Cf. Bechara (1960: 156), Luft (1986: 141) e Cegalla (1996: 252), por exemplo.

Com uma regra gramatical que dê conta desse tipo de estrutura, estamos mostrando formalmente em nossa gramática que o núcleo **P** pode exigir ou não complemento (como discutimos acima) e, além disso, pode **selecionar** a categoria sintática de seu complemento, ou seja, pode **c-selecionar** seu complemento.

Com base nas regras gramaticais que propusemos e com base em nossas análises sobre o comportamento das preposições no sintagma preposicional, chegamos a algumas generalizações empíricas interessantes:

i. apesar de existirem preposições **unicamente transitivas** (*João falou com a Maria* / **João falou com*), não existem preposições **unicamente intransitivas** (*Maria sentou-se atrás* / *Maria sentou-se atrás de mim*). Ou seja: há dois tipos de preposições: as que **exigem** complementos e as que **aceitam** complementos;

ii. o complemento das preposições **unicamente transitivas** (aquelas que exigem complementos) podem ser DPs, CPs, PPs ou AdvPs (*João foi para* $_{DP}$[*o Chile*] / *João saiu para* $_{CP}$[*que pudesse chegar cedo ao trabalho*] / *João foi para* $_{PP}$[*fora de casa*] / *João foi para* $_{AdvP}$[*lá*]);

iii. o complemento das preposições que podem ser **transitivas ou intransitivas** (aquelas que aceitam complementos) deve necessariamente ser um PP (*Maria sentou-se atrás do João* / **Maria sentou-se atrás o João*)[97].

No entanto, devemos admitir aqui que a observação (iii) parece nos levar à ideia de que, em português, não existam preposições intransitivas. Esse grupo de preposições (que aceita complementos mas não os exige) se comporta um pouco como os advérbios: tal como os advérbios, e ao contrário das outras preposições, elas não atribuem Caso estrutural ao seu complemento. Daí termos **[atrás do João]** e não ***[atrás o João]**[98]. Além disso, todas essas preposições exigem como complemento um PP cujo núcleo é a preposição **de** funcional, marca típica de PP complemento de palavra nominal[99]. Deixamos registradas aqui essas duas ressalvas, mesmo que estejamos aceitando os argumentos em favor de uma análise que reconheça preposições transitivas e intransitivas, como vimos[100].

Passemos à análise do sintagma adverbial.

4.4 O SINTAGMA ADVERBIAL

Já falamos sobre o advérbio em algumas passagens anteriores. Agora, iremos apresentar nossas análises sobre a estrutura interna do AdvP. Elaboramos as seguintes regras categoriais de reescrita para a análise do sintagma adverbial:

(4.198) AdvP → Adv'

(4.199) Adv' → AdvP Adv'

(4.200) Adv' → Adv

[97] Apenas à exceção das preposições **sem** e **contra**, que podem aparecer sem complemento quando é possível recuperá-lo contextualmente (ver exemplos da nota 95).

[98] É uma característica dos elementos [+N] não atribuírem Caso estrutural a seus complementos.

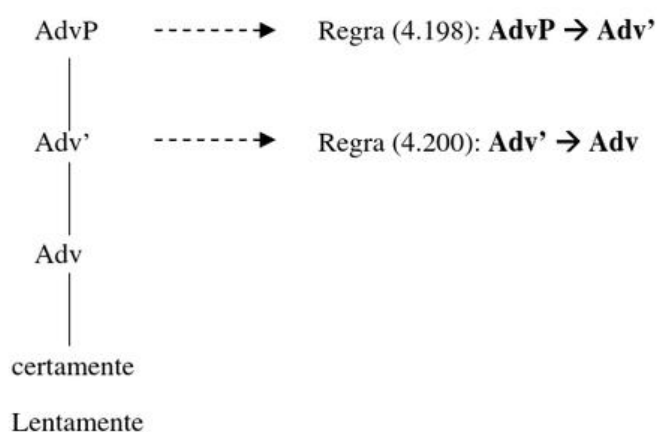
[99] Agradecemos Sérgio Menuzzi, em comunicação pessoal, por ter nos alertado a respeito desse último ponto.

[100] Remetemos o leitor (curiosa e ironicamente) a Othero (2006) para argumentos a favor da proposta de se classificar as nossas preposições intransitivas como advérbios e não como preposições.

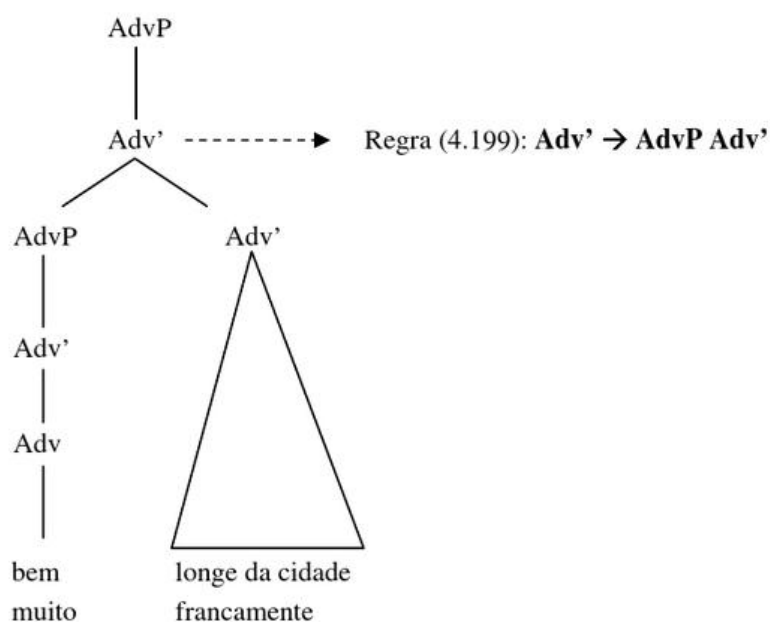
(4.201) Adv' → Adv PP

Vejamos a motivação e a aplicação dessas regras de descrição do AdvP. A primeira regra dá conta da análise de um advérbio intransitivo sem qualquer modificador.

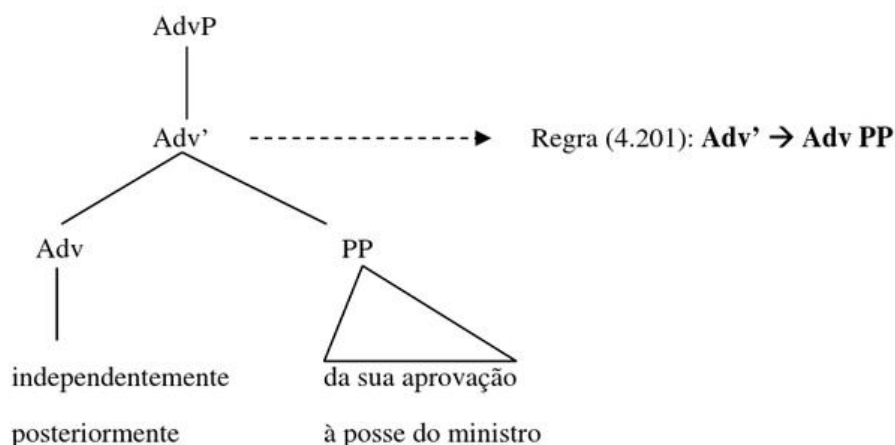
(4.202) O João [certamente] escreveu bastante.

(4.203) [Lentamente], todos se acalmaram.

O AdvP também pode ser modificado por um outro AdvP.

(4.204) Meus amigos moram [bem longe da cidade].

(4.205) João sempre fala [muito francamente].

Também propusemos uma regra que prevê uma estrutura de um AdvP cujo núcleo seja um advérbio transitivo, com um complemento PP[101]. É a regra (4.201).

(4.206) Isso será feito [independentemente da sua aprovação].

(4.207) Os guardas poderão sair [posteriormente à posse do ministro].

As regras parecem adequadas, mas evitamos entrar em um assunto delicado: o que exatamente é um advérbio? Provavelmente, não teríamos fôlego para discutir neste capítulo a natureza de um advérbio. E provavelmente não conseguiríamos chegar a uma conclusão definitiva, uma vez que as evidências empíricas parecem apontar para o fato de que há mais do que uma classe de palavras no que tradicionalmente se classifica sob o rótulo "advérbio"[102]. Remetemos o leitor a Costa & Costa (2001), Perini (2000, 2006) e Saraiva (2008) para uma introdução à problematização sobre a classificação dos advérbios em língua portuguesa. Ernst (2001) discute elementos que atuam com adjuntos na estrutura sintagmática da sentença e traz apontamentos interessantes a respeito dos advérbios.

Passemos finalmente à análise do sintagma verbal.

---

[101] Seguindo Jackendoff (1977), Huddleston (1984) e Lemle (1984), por exemplo. Todos os núcleos lexicais que estudamos até aqui (N, A, P, Adv) podem ser transitivos ou intransitivos.

[102] Já trouxemos duas citações que tratam da complexidade dessa classe que chamamos "advérbio": para Winograd (1983: 53), "[o advérbio] é a sobra entre as categorias (*leftover category*) – qualquer coisa que não se encaixa em outra classe recebe o nome de advérbio"; para Perini (2006: 163-4), "chamar a todos [os advérbios tal como se apresentam na NGB] de 'advérbios' é ignorar um mundo de fenômenos gramaticais importantes".

4.5 O SINTAGMA VERBAL

Nesta seção, apresentaremos as regras para a boa formação do sintagma verbal em português. Alguns pontos envolvendo os VPs ficarão para a próxima seção, quando trabalharemos com a estrutura da sentença e com estruturas com "locuções verbais". Comecemos a estudar a estrutura interna do sintagma verbal em português analisando as seguintes regras:

(4.208) VP → V'
(4.209) V' → AdvP V'
(4.210) V' → V' AdvP
(4.211) V' → V
(4.212) V' → V DP
(4.213) V' → V PP
(4.214) V' → V CP
(4.215) V' → V AP
(4.216) V' → V AdvP
(4.217) V' → V' PP

Essas regras que propusemos se beneficiaram parcialmente das análises de Lemle (1984), Lobato (1986) e Othero (2006). As regras que esses autores apresentaram na tentativa de descrição do sintagma verbal em português podem ser vistas no quadro 4.1, com exceção das regras de Othero (2006), que já apresentamos no capítulo 3:

Regra do sintagma verbal (SV) de Lemle (1984: 171)[103]:

$$SV \rightarrow (Adv)^* \ V \ (Adv) \begin{bmatrix} SN \\ Adj \end{bmatrix} (Sprep)^* \ (S)^*$$

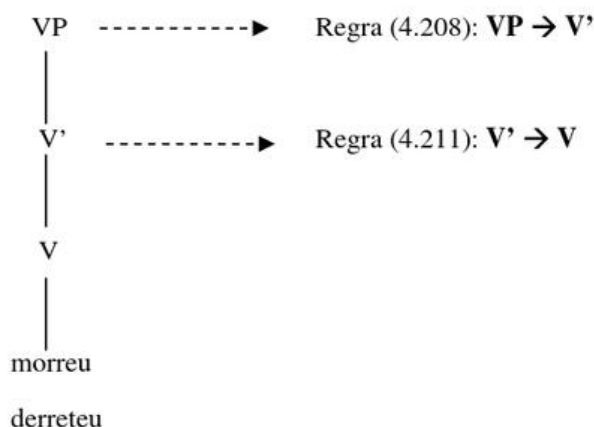
Regra do sintagma verbal (SV) de Lobato (1986: 121)[104]:

$$SV \rightarrow \begin{Bmatrix} V \ (SN) \ (SP)^n \\ Cop \ SA \end{Bmatrix} (SAdv)$$

Quadro 4.1: Regras de Lemle (1984) e Lobato (1986)

Como mostram Lemle (1984) e Lobato (1986), mais uma vez percebemos que o único elemento obrigatório no sintagma é o elemento nuclear, aqui o V. Vejamos algumas análises com as regras que propusemos. Comecemos justamente com estrutura de um VP cujo núcleo é um verbo intransitivo (que não apresenta, portanto, complementos) sem elementos modificadores.

(4.218) A Maria [morreu].

(4.219) Meu sorvete [derreteu].

Essas regras preveem uma estrutura com verbo intransitivo, sem qualquer complemento ou modificador.

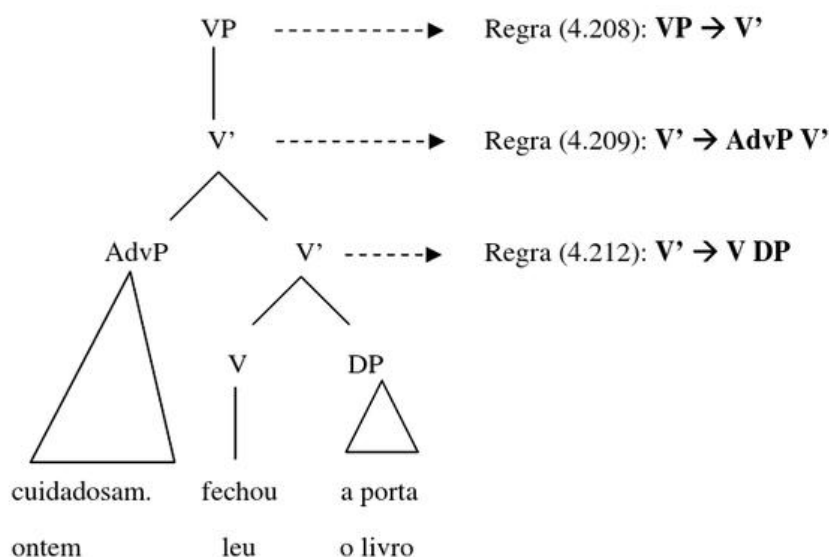
As regras (4.209) e (4.210) dão conta da análise de sintagmas verbais com

[103] Os parênteses indicam a opcionalidade do elemento; o asterisco indica a recursividade dos elementos.

[104] O índice *n* sobrescrito tem o mesmo significado do asterisco; **Cop** representa um verbo copular, ou de ligação; **SA** é o nosso **AP**.

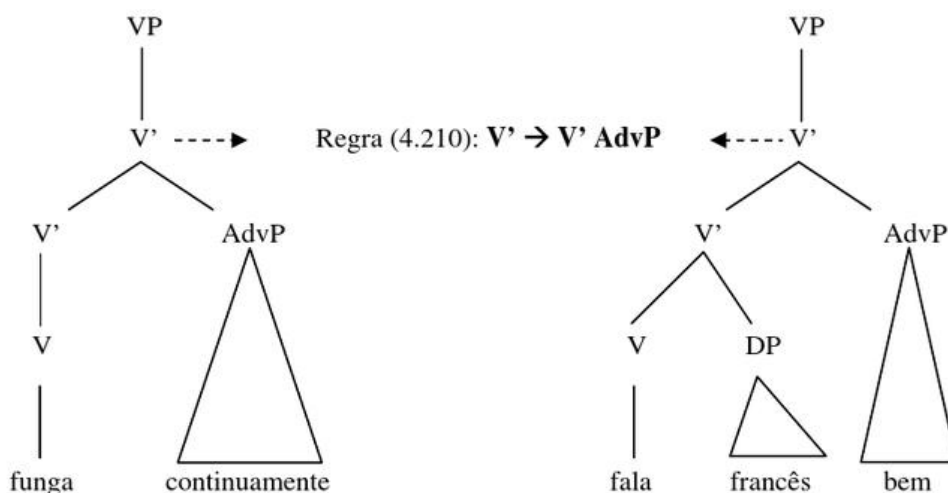
modificação adverbial. Dada a relativa **mobilidade sintática** dos elementos adverbiais no VP[105], elaboramos duas regras para o sintagma verbal modificado por sintagma adverbial: uma em que vemos um AdvP modificando um VP à esquerda de V'; e outra regra que prevê um AdvP modificando um VP à direita de V'. Vejamos algumas análises.

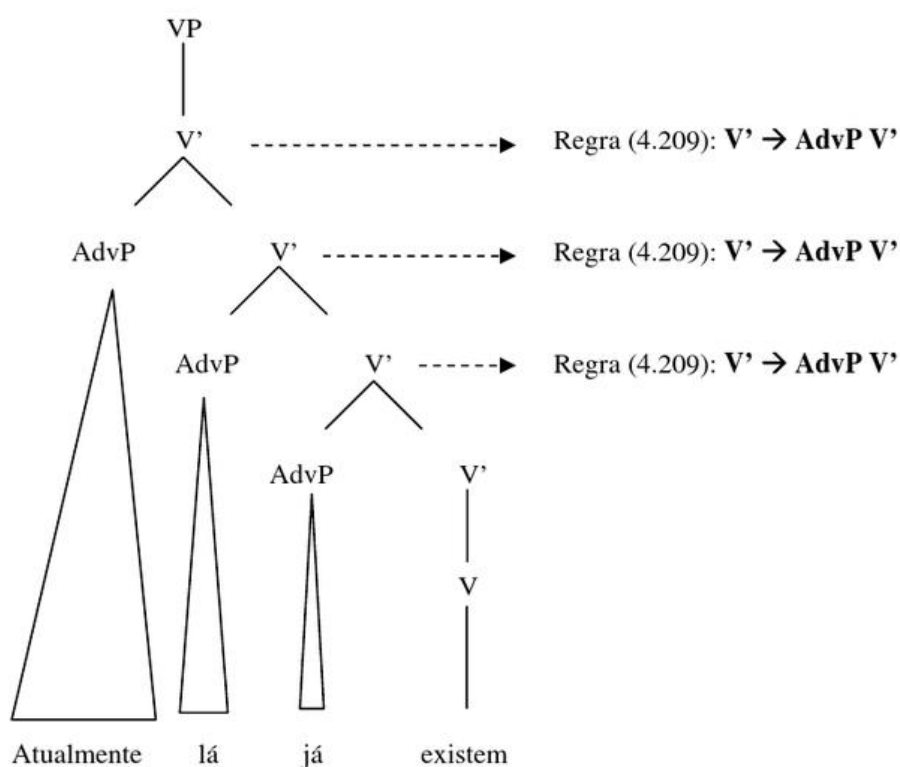
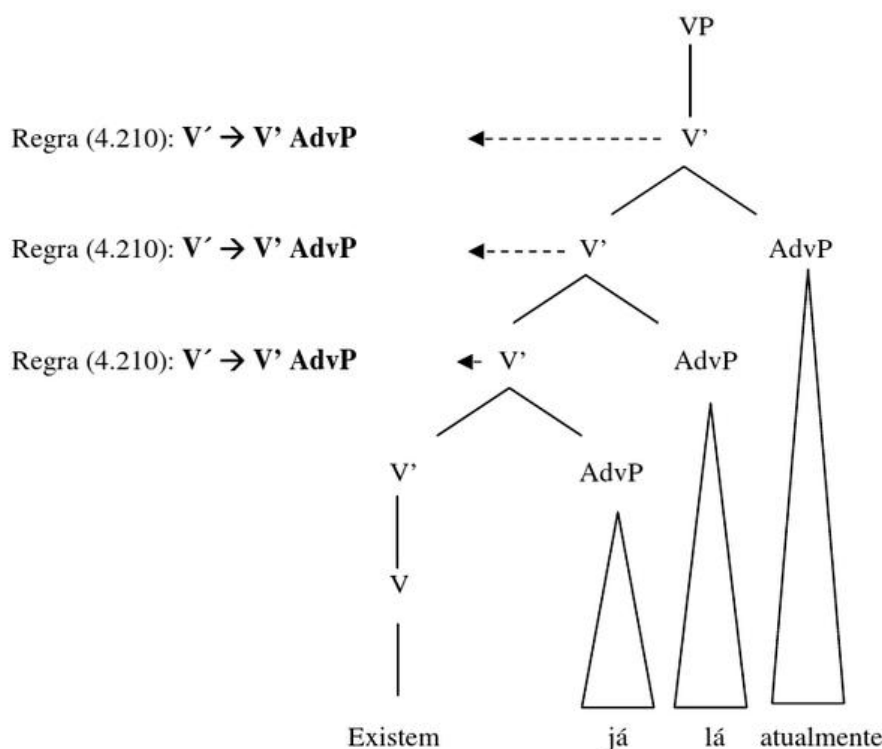
(4.220) O João [cuidadosamente fechou a porta].

(4.221) O João [ontem leu o livro].

Os exemplos acima apresentaram um sintagma verbal com modificador adverbial à esquerda de V'. Abaixo, vemos exemplos com um AdvP modificando um VP à direita de V', tanto um verbo com complemento como um verbo intransitivo.

(4.222) Minha esposa [funga continuamente].

(4.223) João [fala francês bem].

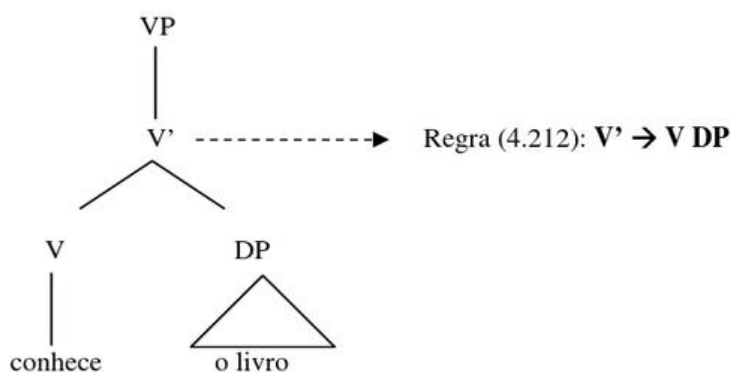
[105] Sobre a posição dos advérbios modificadores no sintagma verbal e na sentença, cf. Jackendoff (1972), Ilari et al. (1990), Cinque (1999), Ernst (2001) e o capítulo 4 de Costa & Costa (2001), além de diversos capítulos publicados no volume II da *Gramática do português falado* (Ilari, 2002). Evidentemente, há mais do que fatores sintáticos atuando na posição dos adjuntos (especialmente de advérbios) no sintagma verbal. Costa (1998), Menuzzi & Mioto (2006) e Figueiredo Silva & Araújo (2008) trabalham com fatores prosódicos e pragmático-informacionais que atuam sobre a determinação da posição do advérbio na sentença em português. Aqui iremos nos restringir a análises estritamente sintáticas. Voltaremos ao assunto adiante.

Um efeito interessante de nossas regras é que elas permitem a recursividade de modificadores adverbiais dentro do VP. Repare nas seguintes análises:

(4.224) [Atualmente lá já existem] mais recursos.

(4.224) [Existem já lá atualmente] mais recursos.

Passemos às regras que fazem referência a verbos transitivos. A primeira regra gramatical que apresentaremos será a regra que analisa sintagmas verbais com um verbo tradicionalmente classificado como **transitivo direto**. Esses verbos exigem ou **aceitam livremente**, para utilizar nomenclatura sugerida por Perini (2000), um complemento DP. Sendo macrorregras de boa formação sintagmática, nossas regras mais uma vez preveem a existência de sintagmas verbais com verbos que **exigem** ou **aceitam** complementos verbais e sintagmas verbais com verbos que **recusam** qualquer complemento (como vimos nas análises de verbos intransitivos acima). As restrições de seleção categorial de cada verbo devem ser implementadas diretamente nos itens lexicais, através de traços lexicais. Perini (2000: 161-173) rediscute a transitividade verbal proposta pela NGB e pelos manuais de gramática e, em seguida, propõe onze subclasses em que os verbos poderiam ser classificados de acordo com a transitividade verbal[106]:

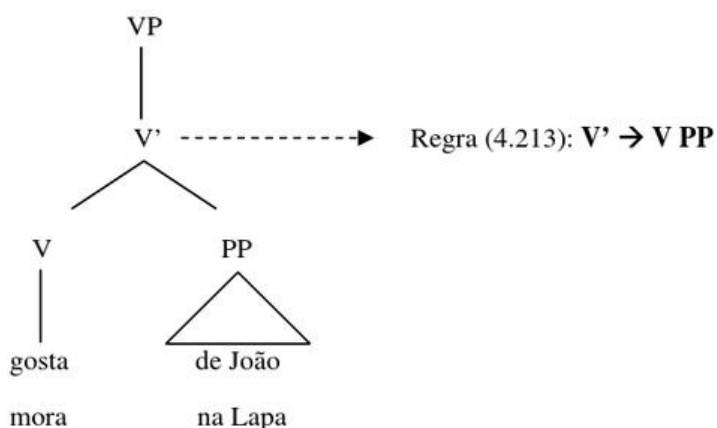
> A gramática tradicional considera relevantes as funções de objeto direto, objeto indireto e predicativo do sujeito (...). O resultado é que tradicionalmente se distinguem cinco subclasses de verbos (transitivos diretos, transitivos indiretos, transitivos diretos e indiretos, intransitivos e de ligação), ao passo que em nosso sistema há onze (...). (Perini, 2000: 166-7)

[106] Discussão semelhante é encontrada também em Azeredo (2000); porém, sem a sistematização que Perini propõe (cf. Azeredo, 2000: 68-86).

Acreditamos que essa classificação pode ser implementada diretamente nas restrições selecionais de cada verbo, enquanto item lexical, ou seja, restringindo as exigências de cada verbo a traços lexicais. Adotando o esquema de Perini, poderíamos implementar onze diferentes traços lexicais aos verbos, de acordo com suas restrições selecionais no que diz respeito à sua complementação (cf. discussão em Perini, 2000: 159-173). As regras de reescrita de nossa gramática irão prever a boa formação de VPs com complemento e sem complemento, enquanto o traço lexical de cada verbo especificará formalmente sua seleção categorial. Vejamos mais exemplos de aplicações de nossas regras.

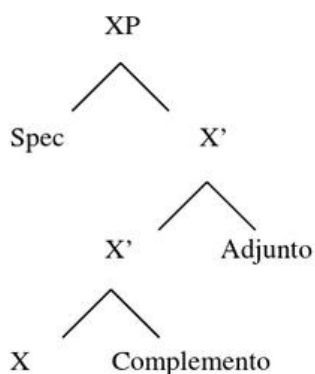
(4.225) Ninguém [conhece o livro].

A regra (4.213) analisa verbos cujo complemento é um sintagma preposicional, tanto verbos tradicionalmente classificados como **transitivos indiretos** como verbos que aceitam ou exigem **complemento circunstancial** (Câmara Jr., 1981; Rocha Lima, 1984; Luft, 1986).

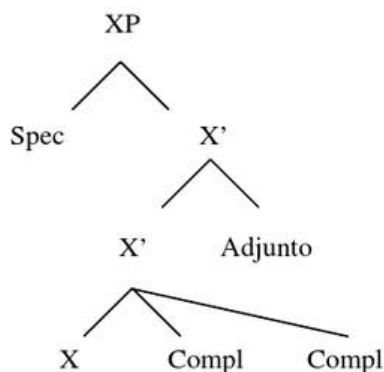
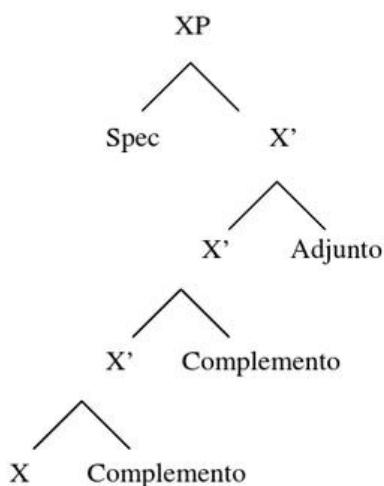
(4.226) Maria [gosta de João].

(4.227) João [mora na Lapa].

Há verbos que aceitam dois complementos, um DP e um PP:

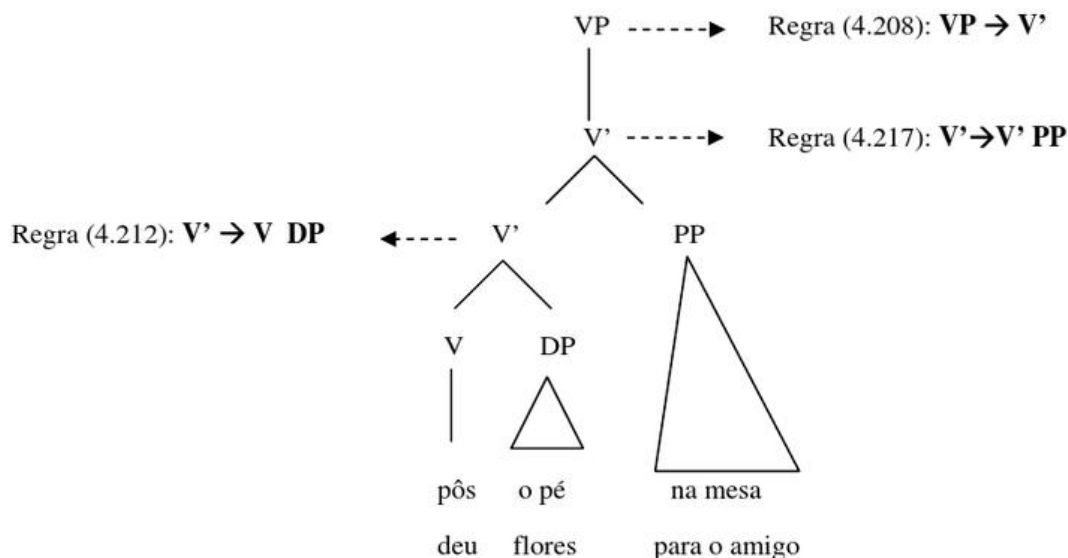
(4.228) João [pôs o pé na mesa].

(4.229) Maria [deu flores para o amigo].

Antes de apresentarmos nossa representação arbórea dessas sentenças com verbos que apresentam dois complementos, vejamos novamente o modelo da teoria X-barra que estamos utilizando:

O modelo X-barra prevê a posição de apenas **um** complemento do núcleo, o que parece ser um problema para representar estruturas em que um núcleo exija ou aceite dois complementos, como agora, com as análises dos verbos ditos **bitransitivos**. Para representarmos estruturas cujo núcleo exige mais de um complemento, deveremos sacrificar o princípio de **binaridade** (cf. capítulo 2) ou postular um nível a mais na representação sintática. Em outras palavras, deveremos adotar uma das duas seguintes representações estruturais: (4.230), que mantém os complementos irmão do núcleo; ou

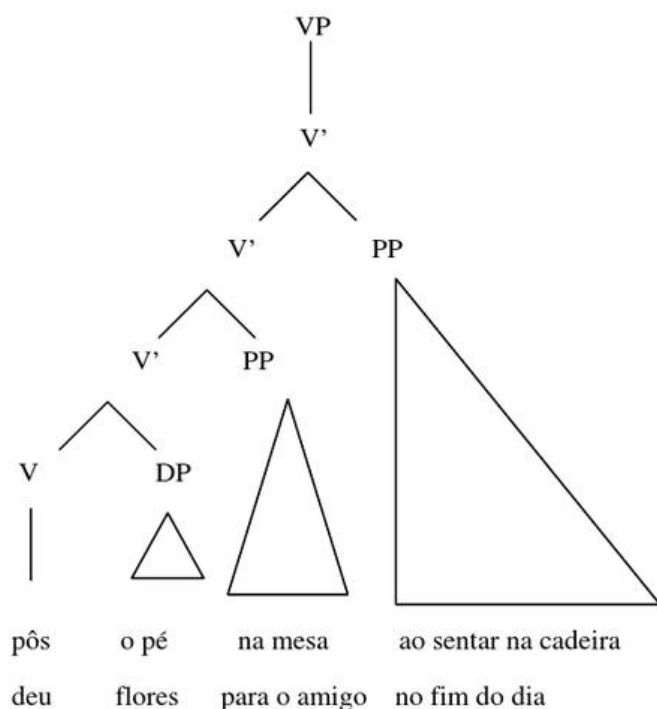
(4.231), que respeita binaridade em uma construção cujo núcleo apresenta dois complementos, mas postula um nível a mais da estrutura intermediária X'.

Por motivos internos à teoria, para manter a estrutura binária, mas também seguindo a tendência geral em trabalhos de descrição sintática que utilizam a teoria X-barra (Haegeman, 1995; Culicover, 1997; Carnie, 2002; Mioto et al., 2004, entre muitos outros), iremos representar construções com dois complementos – como nos casos dos verbos bitransitivos – da maneira como apresentamos em (4.231). Respeitaremos, portanto, o princípio da binaridade. Essa é, na verdade, a maneira tradicional de lidar com o problema de núcleos bitransitivos dentro do modelo da GB e P&P. Dessa forma, mantemos a estrutura binária, mesmo que tenhamos um "'desencontro' entre sintaxe-semântica, uma vez que um argumento semântico acaba ocupando uma posição que não é

a de argumento sintático[107]", como dizem Santorini & Kroch (2007: 70).

Portanto, a representação da estrutura sintagmática de sentenças como (4.228) e (4.229) será a seguinte:

Repare que não precisamos elaborar nenhuma regra além daquelas que já havíamos apresentado até agora. O PP complemento está ocupando uma posição de adjunto no esquema da X-barra. Como a regra é recursiva, podemos ter também um VP com dois complementos e com um adjunto preposicional, como nas sentenças abaixo:

(4.230) Maria [pôs o pé na mesa ao sentar na cadeira].

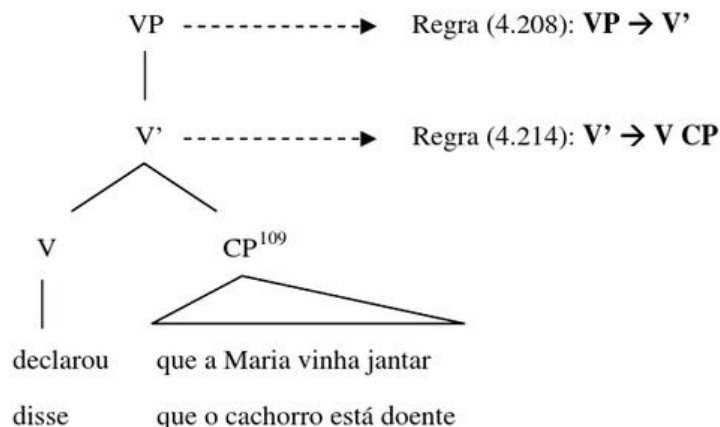
(4.231) Maria [deu flores para o amigo no fim do dia].

[107] Trecho original: "(...) syntax-semantics mismatch, since a semantic argument ends up occupying a position that is not a syntactic argument position". Na verdade, há ainda uma terceira maneira de lidar com verbos que exigem dois argumentos, adotando a hipótese apresentada por Larson (1988, 1990), os **VP shells**. Não seguiremos essa ideia aqui porque acreditamos que ela torna a estrutura sintática complexa demais.

As regras parecem estar adequadas o suficiente para analisarmos VPs com essas estruturas (sem complementos, com um complemento, com dois complementos – todos com ou sem a presença de adjuntos). No entanto, por termos estabelecido que um PP complemento ocupa a mesma posição na estrutura arbórea que um PP adjunto, não há nenhum lugar em nossas regras que previna a má formação de um VP com dois, três ou mais complementos preposicionais, o que é raro de acontecer[108]. Daí a importância de um trabalho com traços lexicais, que levem a informação de complementação categorial de cada item lexical para o componente das regras de boa formação da sentença.

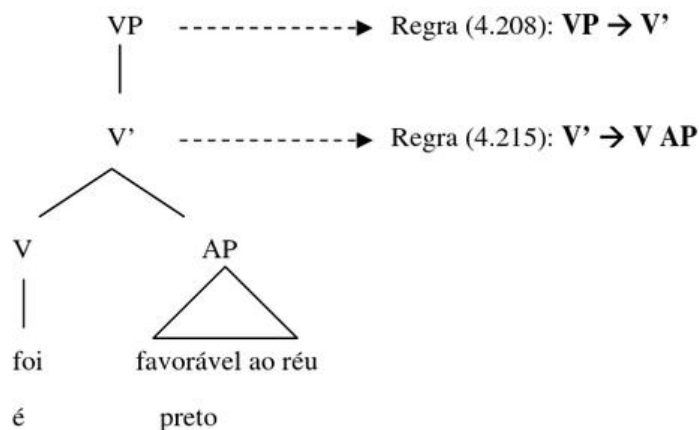
Passemos às demais regras gramaticais que elaboramos para a análise do sintagma verbal. Um último tipo de complemento verbal (de verbos ditos **plenos**; falaremos sobre os complementos de verbos copulares a seguir) é aquele que pode ser expresso por um elemento oracional, ou seja, um CP.

(4.232) O João [declarou que a Maria vinha jantar].

[108] Encontramos poucos verbos que exijam mais de dois complementos (*apostar*, *viajar*, etc.). O assunto da estrutura argumental é discutido com detalhe em Levin (1993); alguns exemplos são discutidos em Jackendoff (2002, cap. 5). Um exemplo clássico que é corretamente previsto pelas regras de nossa gramática é o seguinte:
a. Maria $_{VP}$[apostou $_{DP}$[cem reais] $_{PP}$[com o João] $_{PP}$[sobre o jogo de ontem]].
Além do mais, nossas regras conseguem analisar corretamente verbos que exigem dois complementos preposicionados, como o verbo *falar*, em (b):
b. João $_{VP}$[falou $_{PP}$[de seu trabalho] $_{PP}$[para a Maria]].

(4.233) O veterinário [disse que o cachorro está doente].

Por fim, vejamos as duas últimas regras de descrição do sintagma verbal em PB. Seguindo a tradição dos estudos em sintaxe gerativa (Chomsky, 1981, 1986b; Radford, 1981, 1988, 1997a,b; Lemle, 1984; Lobato, 1986; etc.) consideraremos os "verbos de ligação" (ou verbos copulares) como núcleos do VP[110]. O complemento de um verbo copular pode vir expresso por um sintagma adjetival ou adverbial[111]:

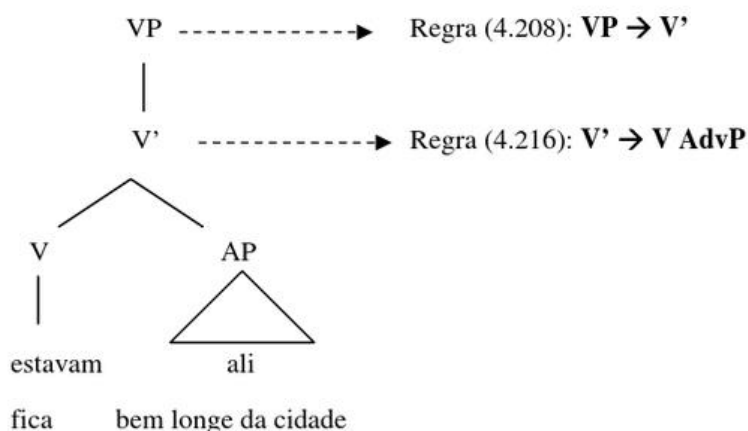
(4.234) A decisão [foi favorável ao réu].

(4.235) O carro de João [é preto].

[109] Estudaremos a estrutura interna do CP na próxima seção, em 4.6.

[110] Ao contrário da posição que adotamos aqui, há quem classifique os verbos copulares **ser** e **estar** como núcleos de IP e não do VP (cf. Becker, 2000, por exemplo). Na seção seguinte, veremos que alguns verbos podem também ser núcleos de IP, atuando como verbos auxiliares.

[111] Na verdade, os complementos de verbos copulares também podem ser DPs (*O João [é [um bom amigo]]*) ou PPs (*O sorvete [é [de creme]]*). Esses casos podem ser analisados com as regras do VP que já apresentamos: (4.212) e (4.213).

(4.236) Os livros [estavam ali].

(4.237) A casa de Alice [fica bem longe da cidade].

Antes de terminar, devemos fazer um *mea culpa*: nossas regras dão conta da estrutura interna do sintagma verbal apenas em sua ordem não marcada, ou **canônica**, em português: **Verbo + Complemento + Advérbios**. Elaboramos nossas regras gramaticais sem prever um componente transformacional que possibilitasse a análise de constituintes deslocados[112]. Por isso, as regras gramaticais que apresentamos aqui podem descrever apenas estruturas não marcadas, que sigam a ordem canônica, em PB.

A ordem dos constituintes em português é relativamente livre, e existem mecanismos bastante específicos que "autorizam" a inversão em sua ordem dentro do sintagma verbal e dentro da sentença. Já havíamos atentado para o fato de que há mais do que fatores sintáticos estritos atuando sobre a ordem dos constituintes no sintagma verbal. Ilari (1992), Pezzati & Camacho (1997), Costa (1998) e Menuzzi (2004), por exemplo, mostram como fatores pragmáticos, discursivos e relativos à estrutura informacional da sentença atuam na formação de estruturas marcadas no português. Menuzzi & Mioto (2006) e Figueiredo Silva & Araújo (2008) apontam para os fatores prosódicos e informacionais que atuam sobre a determinação da posição do advérbio na sentença e no VP em português.

Se estipulássemos regras de reescrita que descrevessem estruturas com constituintes deslocados, essas regras iriam deixar escapar generalizações interessantes (como em quais contextos um constituinte pode ser deslocado, ou que tipo de constituinte pode ser deslocado). Além do mais, essas novas regras figurariam, em nossa gramática, no

[112] Na verdade, vimos também a estrutura **Adv V (Complemento)**.

mesmo nível hierárquico que as regras de descrição da ordem não marcada da sentença, o que implicitamente estaria sugerindo que um elemento X pode, irrestritamente, aparecer tanto na posição depois do núcleo ou antes do núcleo, depois do elemento Y ou antes dele, etc. Em outras palavras, uma gramática que contivesse todas as regras de descrição da estrutura superficial de todas as ordens entre os elementos na sentença em português deixaria escapar a ideia de que o português apresenta uma estrutura básica de ordem dos constituintes na frase, a **ordem direta**, **não marcada**: Sujeito – Verbo – Complemento(s) – Adjunto(s); e que essa ordem pode ser violada por algum motivo de outra natureza gramatical, como por fatores prosódicos, informacionais, pragmáticos ou de processamento.

A seguir, passaremos ao estudo da estrutura da sentença em PB. Também analisaremos apenas sentenças na ordem direta, SVO + Adjunto(s).

4.6 ANÁLISE DA SENTENÇA EM PB

Finalmente, nesta última seção, iremos apresentar as regras gramaticais que descrevem a estrutura da sentença em PB. Trabalharemos com a estrutura da sentença declarativa simples em sua ordem canônica (ou direta), com sentenças encaixadas e com sentenças com verbos auxiliares. Começaremos por aquelas sentenças que apresentam "locuções", ou agrupamentos verbais, – construções com **verbos auxiliares + verbos principais**. Baseamo-nos principalmente em dois trabalhos pioneiros que se debruçaram sobre as construções de "locuções verbais", Pontes (1973) e Lobato (1975)[113].

[113] Esses dois trabalhos servem ainda hoje como referência no estudo dos auxiliares e das locuções verbais em português. Pontes (1973) é a versão publicada de sua tese de livre docência em Linguística, defendida em 30 de abril de 1971, na Universidade Federal de Minas Gerais. Lobato (1975) é a versão publicada (e estendida) de um capítulo da tese de Doutorado de autora, defendida em fevereiro de 1971 na Université de Paris III (o título de sua tese é *L´auxiliarité en langue portugaise*).

4.6.1 Sentenças com verbos auxiliares

Pontes (1973) e Lobato (1975) se debruçaram sobre o assunto das "locuções verbais"[114] para tentar conceituar o que realmente é uma construção com um verbo auxiliar, distinguindo-as de outras estruturas que apresentam uma sequência de dois ou mais verbos.

Apesar dos numerosos estudos sobre as "orações complexas" em português desde os trabalhos pioneiros de Pontes (1973) e Lobato (1975), a tradição gramatical não parece ter se beneficiado dos resultados obtidos por esses trabalhos, como fica fácil de perceber se consultarmos algumas gramáticas modernas de língua portuguesa, que ainda trazem nomenclatura confusa quando lidam com termos como *locução verbal*, *oração*, *verbo principal* e *verbo auxiliar*. Gonçalves (1996: 7), que analisou 49 construções verbais compostas comumente relacionadas em gramáticas normativas da língua portuguesa, relata que

> existem em Português e noutras línguas construções em que ocorre uma sequência de verbos que, nalguns manuais de análise linguística elaborados no quadro da Gramática Tradicional, é designada por locução verbal e definida como "(...) a combinação das diversas formas de um verbo auxiliar com o infinitivo, gerúndio ou particípio de outro verbo que se chama principal (...)." (Bechara, 1963: 110)
> Assume-se, assim, a existência de duas grandes classes de verbos: os auxiliares (Vaux) e os principais (Vprinc). Sendo esta distinção unanimemente aceite, seria de esperar que a delimitação dos membros de cada classe fosse também unânime. No entanto, a partir da leitura de algumas das chamadas Gramáticas Tradicionais, é difícil chegar a uma conclusão sobre quais os verbos que pertencem a cada uma das classes acima mencionadas (...)". (grifos da autora)

De qualquer maneira, esse é um assunto que não levaremos adiante aqui. Queremos apenas mostrar que as construções chamadas tradicionalmente de "locuções verbais", na verdade, servem de rótulo para ao menos dois tipos distintos de estruturas: (i) uma estrutura com um verbo auxiliar na posição de **Infl** e um verbo principal, núcleo de VP; e (ii) uma estrutura com duas orações distintas, ou seja, com dois verbos principais que ocupam a posição nuclear de dois VP distintos. Iremos analisar sentenças como as seguintes:

[114] Chamaremos de "locuções verbais" toda a sequência de verbo auxiliar + verbo principal. Mas manteremos o termo entre aspas por sabermos que há muita confusão terminológica envolvendo o rótulo "locuções verbais".

(4.238) João vai gostar do filme.

(4.239) João parece gostar do filme.

Apesar de terem estruturas aparentemente idênticas **[DP Vflexionado Vinfinitivo PP]**, essas sentenças são estruturalmente distintas. Em (4.238), temos um verbo auxiliar seguido de um verbo principal; em (4.239), temos dois verbos principais. Sabemos disso porque podemos aplicar alguns "testes de auxiliaridade" que nos indicam se estamos lidando com uma construção que apresenta dois verbos plenos ou com uma estrutura de verbo auxiliar mais verbo pleno. Por exemplo, um verbo auxiliar não costuma permitir a negação de seu verbo principal. Em uma construção com dois verbos plenos, ao contrário, podemos negar tanto o primeiro como o segundo verbo.

(4.240) a. O João vai gostar do filme.
b. O João **não** vai gostar do filme.
c. *O João vai **não** gostar do filme.

(4.241) a. O João parece gostar do filme.
b. O João **não** parece gostar do filme.
c. O João parece **não** gostar do filme.
d. O João **não** parece **não** gostar do filme.

Isso acontece porque o auxiliar não tem conteúdo proposicional que possa ser negado. Podemos ter apenas **uma** negação por agrupamento verbal[115].

Outro teste de auxiliaridade que podemos aplicar envolve elementos modificadores de tempo. Se estivermos frente a uma estrutura com dois verbos plenos, é possível modificar cada um deles com um modificador temporal distinto, sem tornar a frase agramatical. Entretanto, se estivermos diante de uma construção com verbo auxiliar, fica impossível modificar cada um dos verbos com modificadores temporais diferentes.

(4.242) a. João tinha ido à praia ontem.
b. *Na semana passada, João tinha ido à praia ontem.

[115] Evidentemente, em alguns dialetos do PB, podemos ter mais de um elemento negativo na frase, mas apenas **uma** negação verbal atuando sobre o conteúdo proposicional, como no seguinte exemplo:
A: Você viu Maria por aí?
B: Não, não vi ela não.

(4.243) a. João queria ir à praia depois de amanhã.

b. Ontem à noite, João queria ir à praia depois de amanhã.

Cada verbo principal aceita uma modificação temporal distinta, enquanto o verbo auxiliar, por lhe faltar conteúdo proposicional, falha ao receber modificação temporal.

Um último teste comumente aplicado para identificarmos os verbos auxiliares consiste no teste de s-seleção do sujeito verbal. Um verbo auxiliar, sendo um **núcleo funcional**, não apresenta restrições semânticas em relação a seu argumento externo, ou seja, em relação ao sujeito da sentença. Um verbo pleno, ao contrário, justamente por ser um **núcleo lexical**, s-seleciona seu argumento externo:

(4.244) a. A Maria vai cair!

b. A Bovespa vai cair!

c. Minha gramática vai cair!

(4.245) a. A Maria odeia cair.

b. ??A Bovespa odeia cair.

c. *Minha gramática odeia cair.

Com base em testes como esses, podemos estabelecer critérios para a definição dos verbos auxiliares em português. Pontes (1973), Lobato (1975) e Gonçalves (1992, 1996), que apresentaram estudos aprofundados das chamadas "locuções verbais" em português, apresentam testes como esses, além de alguns outros, para estabelecer os "critérios de auxiliaridade" para o português[116]. Em PB, parece ser consensual que os seguintes verbos podem ser utilizados com o estatuto de auxiliar: *andar*, *estar*, *haver*, *ir*, *ser*, *ter* e *vir*.

Contudo, saber distinguir um verbo pleno de um verbo auxiliar não será nossa preocupação principal aqui, uma vez que, como o leitor já bem sabe, essa distinção deverá ser caracterizada como um traço no item lexical, algo que foge de nossa alçada. Na verdade, iremos apresentar nossas regras gramaticais que descrevem adequadamente estruturas de sentenças com verbos auxiliares e estruturas de sentenças que apresentam construções com dois verbos plenos. O que nos interessa saber aqui é que os verbos

[116] Cf. também Perini (2000), que aplica um outro teste de auxiliaridade, baseando-se no "conjunto de traços de transitividade" (p. 180) de cada verbo. O teste consiste em saber se os efeitos de transitividade de cada verbo são respeitados na "locução verbal": "Observa-se que qualquer traço de transitividade que valha para um verbo quando pertence a um predicado simples vale igualmente para esse verbo quando é parte de um predicado complexo" (p. 179).

auxiliares atuam como um **elemento funcional** na sentença e não como um elemento **lexical**, que apresenta restrições de seleção semântica e conteúdo proposicional. Por isso, a estrutura de uma sentença com a sequência **verbo auxiliar + verbo lexical** será diferente da estrutura de uma sentença que apresente a sequência **verbo lexical + verbo lexical**[117].

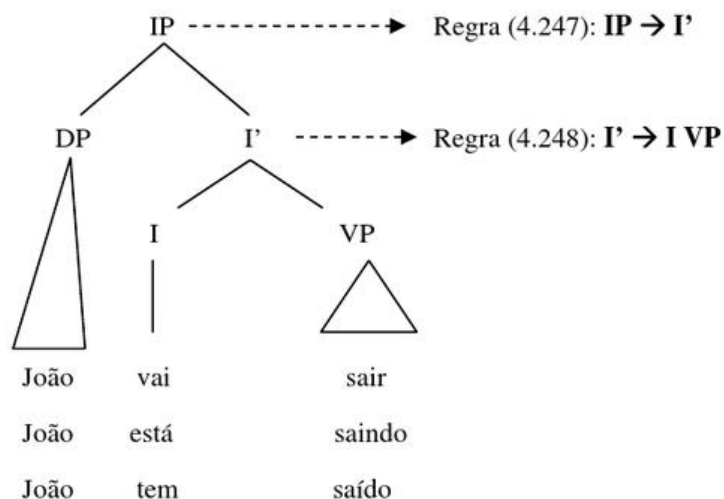
Se estivermos frente a uma construção com dois verbos plenos, teremos duas projeções máximas desses elementos, ou seja, teremos dois VPs. Se estivermos lidando com uma sequência de verbo auxiliar seguido de verbo pleno, o auxiliar será tratado como um núcleo funcional, núcleo do **IP**, como já vimos no capítulo 2 (cf. seção 3 do cap. 2), quando mostramos os núcleos de CP e de IP[118].

Para analisar sentenças com verbos auxiliares, elaboramos as seguintes regras gramaticais:

(4.246) IP → DP I'
(4.247) IP → I'
(4.248) I' → I VP
(4.249) I' → I IP

O núcleo I, como todo núcleo funcional, c-seleciona seu complemento. O complemento de I poderá ser um VP ou um IP. Vejamos a análise de algumas sentenças.

(4.250) João vai sair.
(4.251) João está saindo.
(4.252) João tem saído.

117 Há quem diga – e com certa razão – que a distinção entre verbo auxiliar e verbo principal não é tão clara como tentamos mostrar aqui. Dutra (2003), por exemplo, aponta para um *continuum* entre o **uso** de um verbo como auxiliar ou como verbo pleno. Em uma frase como (a), a "locução verbal" pode ser gradativamente desfeita à medida que inserimos material linguístico entre o verbo auxiliar e o verbo principal:
(a) O João **vai arrumar** um armário quebrado amanhã na casa da Maria.
(b) O João **vai** amanhã **arrumar** um armário quebrado na casa da Maria
(c) O João **vai** amanhã na casa da Maria **arrumar** um armário quebrado.
Em (c), o auxiliar parece ter passado a expressar mais significação própria do que em (a).
118 Como mencionamos no capítulo 2, os verbos auxiliares são considerados núcleo da projeção funcional **I(nf)P** desde Chomsky (1981). Apesar de ser um consenso em literatura sintática dentro do modelo de Princípios e Parâmetros classificar os auxiliares como núcleo de IP, há sintaticistas, como Gonçalves (1992, 1996), que tratam os auxiliares como núcleos do VP. Remetemos o leitor aos trabalhos da autora para ver suas motivações e seus argumentos em favor de sua análise.

O sujeito da sentença, o DP [João], ocupa a posição [Spec, IP], o verbo auxiliar ocupa a posição de núcleo de IP e o verbo (*sair*, *saindo*, *saído*) ocupa a posição de núcleo do VP.

Também é uma construção comum em PB uma locução verbal com dois verbos auxiliares e um verbo principal, como nos exemplos seguintes:

(4.253) Eu vou estar entrando em contato com a senhora.

(4.254) O João tinha estado contando suas piadas.

Essas frases são bem formadas em PB porque a categoria sintática dos complementos dos auxiliares é respeitada. Vimos que, sendo um núcleo funcional, o auxiliar não tem restrições de s-seleção de seus argumentos. No entanto, como todos os núcleos funcionais, o auxiliar apresenta restrições de c-seleção que devem ser respeitadas. Os verbos auxiliares que mostramos anteriormente, por exemplo, se dividem entre três subcategorias de acordo com suas restrições selecionais:

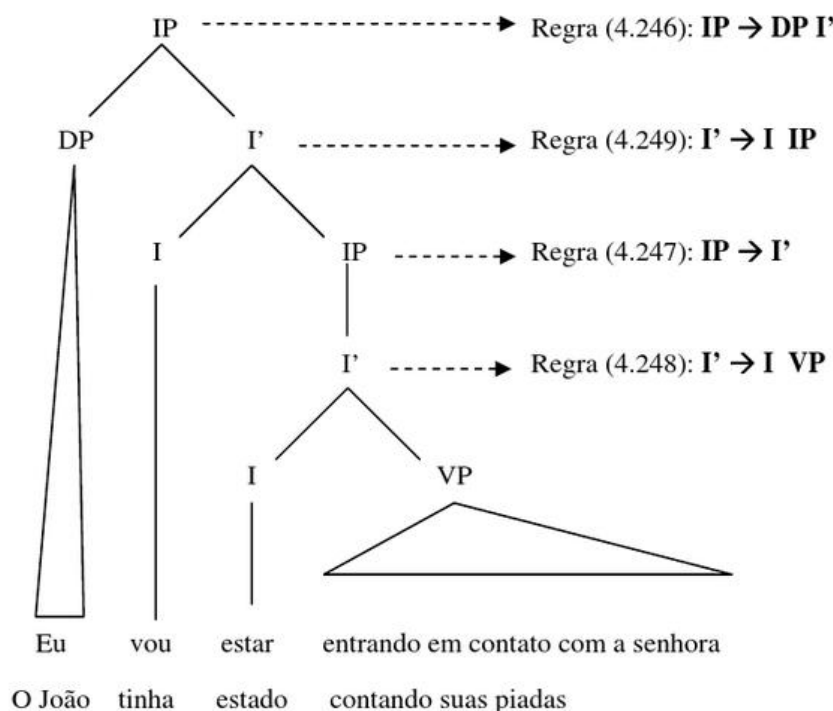
a) *ir* seleciona um VP com V infinitivo;
b) *ser, ter, haver* selecionam um VP com V particípio;
c) *estar, vir, ir, andar* selecionam um VP com V gerúndio.

Além das restrições de seleção de cada auxiliar, há em português uma ordem que deve ser respeitada. De acordo com Perini (2000: 75),

> o predicado complexo é, pois, sempre composto de auxiliar (Aux) mais NdP ["núcleo do predicado", nosso V]; e pode haver mais de um auxiliar, posicionados sempre segundo uma ordenação rígida (...) (primeiro o auxiliar seguido de infinitivo, depois o auxiliar seguido de particípio e, por último, o auxiliar seguido de gerúndio). Se quisermos fazer uma frase utilizando vários auxiliares, teremos de atender às duas exigências acima formuladas: primeiro, a coocorrência de cada tipo de Aux com um de seus verbos auxiliares específicos; e, depois, a ordenação infinitivo – particípio – gerúndio. (Grifos do autor)

As sentenças (4.253) e (4.254) são bem formadas justamente por obedecerem a estes três fatores: (i) obedecem às regras gramaticais de boa formação em PB; (ii) obedecem às restrições de seleção categorial de cada auxiliar; e (iii) obedecem ao ordenamento entre os auxiliares. Vejamos as árvores sintáticas das duas sentenças:

(4.253) Eu vou estar entrando em contato com a senhora.

(4.254) O João tinha estado contando suas piadas.

Essas regras gramaticais nos permitem analisar sentenças com a construção **verbo auxiliar + verbo pleno**[119]. Antes de analisarmos as sentenças com a sequência de dois

[119] Não abordamos a questão de verbos auxiliares que têm complementação não finita não imediata, como alguns modais e aspectuais (*ter que*, *ter de*, *estar a*, *começar a*, etc.). Não está claro para nós qual é a estrutura desses verbos, nem a natureza de seus complementos. Para alguns, a preposição exigida por esse

verbos plenos, passemos à nossa análise da sentença simples em PB, com apenas um verbo principal (e em sua forma canônica).

4.6.2 Sentenças simples

A sentença simples é aquela que apresenta apenas um verbo, em posição de núcleo do VP. Antes de passarmos às regras gramaticais e às nossas análises, vamos discutir brevemente um pouco sobre os **níveis** de representação sintática dentro do quadro teórico da GB.

Dentro do paradigma gerativista e do modelo da GB, as sentenças são comumente analisadas em dois níveis de representação sintática: a **estrutura-P** e a **estrutura-S**. A estrutura-P é a estrutura de organização sintática **subjacente**, que antecede qualquer operação ou derivação sintática; é o nível que "codifica as propriedades lexicais dos constituintes da sentença (...) [, que] representa as relações argumentais básicas na sentença" (Haegeman, 1995: 304). A **estrutura-S** é o nível de representação sintática resultante dos processos de movimento e derivação sintática, é o nível que "reflete as estruturas mais superficiais da sentença; a ordem final dos elementos na seqüência de superfície" (Haegeman, 1995: 305)[120].

Nas análises que levam em conta esses dois níveis de representação sintática, o verbo é considerado núcleo de VP na estrutura-P, ao passo que um **morfema de flexão verbal** ocupa a posição de núcleo de IP, como mostramos com a análise da estrutura-P da seguinte frase[121]:

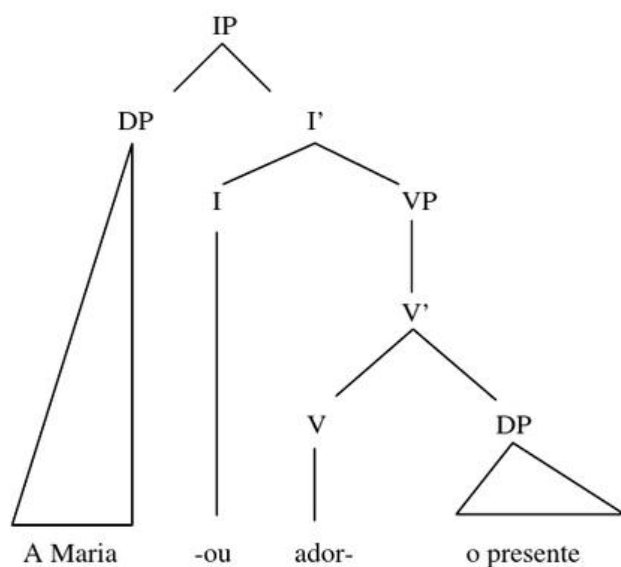
(4.255) A Maria adorou o presente.

---

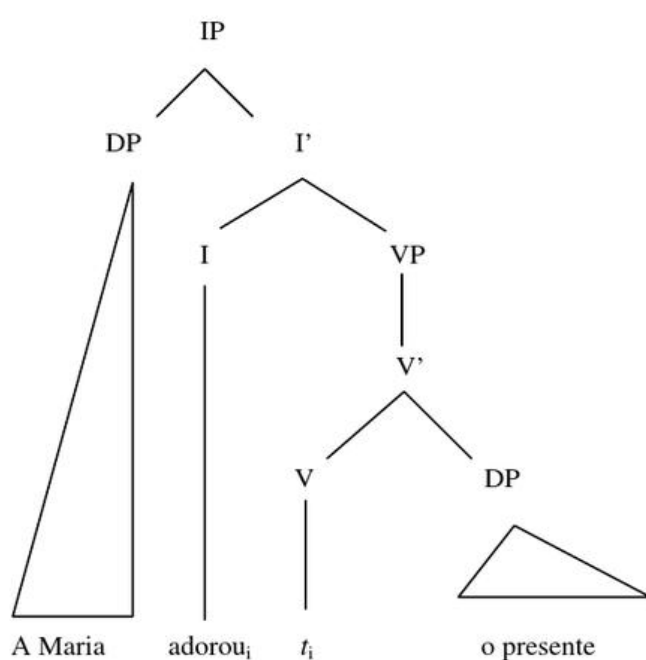
tipo de verbo é parte integrante do verbo (cf. Burzio, 1986, Corso, 2002; Boff, 2003); para outros, ela é um núcleo funcional do complemento do verbo, desempenhando aí diversas funções (cf. Raposo, 1989; Gonçalves, 1996; Lunguinho, 2008a,b, a sair). Remetemos o leitor a essas referências para discussão. Lunguinho (2008b) traz um bom panorama do problema.

[120] Trechos originais: "(...) encodes the lexical properties of the constituents of the sentence (…) represents the basic argument relations in the sentence". Sobre a estrutura-P: "(…) reflects the more superficial properties of the sentence: the actual ordering of the elements in the surface string".

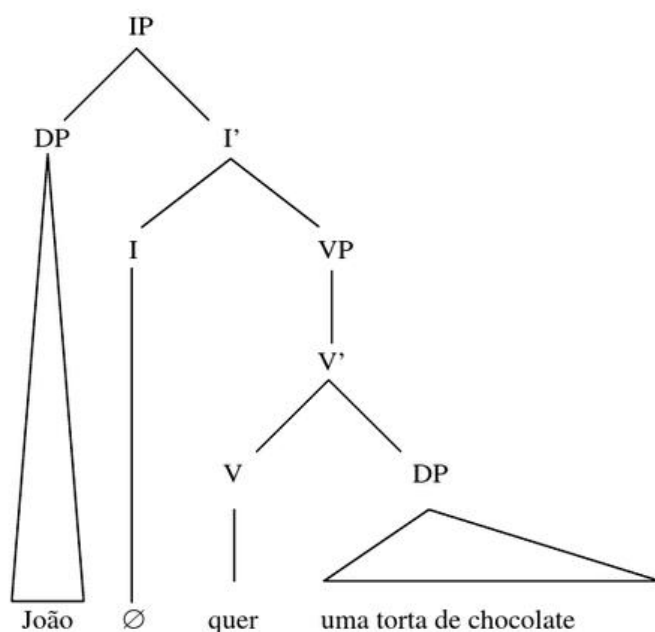
[121] Não discutiremos aqui a divisão de IP em TP e AgrP por não ser importante para o ponto que iremos discutir. Não estamos adotando aqui a hipótese de que o sujeito da sentença surja na posição [Spec,VP] em estrutura-P. Esse ponto é irrelevante para nossa discussão.

Estrutura-P:

A sequência acima**, [a, Maria, -ou, ador-, o presente]**, no entanto, não pode ser pronunciada. Para que seja pronunciada, essa estrutura deve passar por uma derivação: o verbo deve mover-se para I e se unificar com sua flexão. Depois do movimento do verbo para I, teremos a seguinte estrutura:

(4.255) A Maria adorou o presente.

Estrutura-S:

Essa estrutura está mais próxima de sua forma pronunciada do que a anterior. Analisando a sintaxe de superfície da sentença, percebemos que o verbo se moveu de sua posição de núcleo do VP e passou a ocupar, juntamente com seu morfema de flexão, a posição de núcleo de IP. A posição de V, em estrutura-S, é ocupada pelo vestígio do verbo (*t*), coindexado a ele ($_i$). O vestígio é um elemento vazio, foneticamente não pronunciado[122].

Isso acontece, evidentemente, quando há um morfema flexional verbal que esteja ocupando a posição de núcleo de IP. Vimos na seção anterior que, em predicados com a sequência **verbo auxiliar + verbo principal**, quem ocupa a posição de núcleo de IP é o verbo auxiliar. Também acontece que, às vezes, em sentenças simples, o morfema flexional do verbo é um elemento vazio, representado pelo símbolo ∅, como na sentença abaixo:

(4.256) João quer uma torta de chocolate.

[122] Um argumento empírico que mostra que os vestígios de elementos movidos, apesar de não pronunciados, ocupam um lugar na estrutura fonético-fonológica da sentença é o que envolve o *wanna contraction* em inglês (cf. Lakoff, 1970; Chomsky & Lasnik, 1977; Pesetsky, 1982, entre outros). Em inglês, a contração do verbo *want* + *to* para formar *wanna* é comum em língua falada. Contudo, a contração torna-se impossível quando houver um vestígio entre o verbo e a partícula *to*, como acontece no seguinte exemplo:

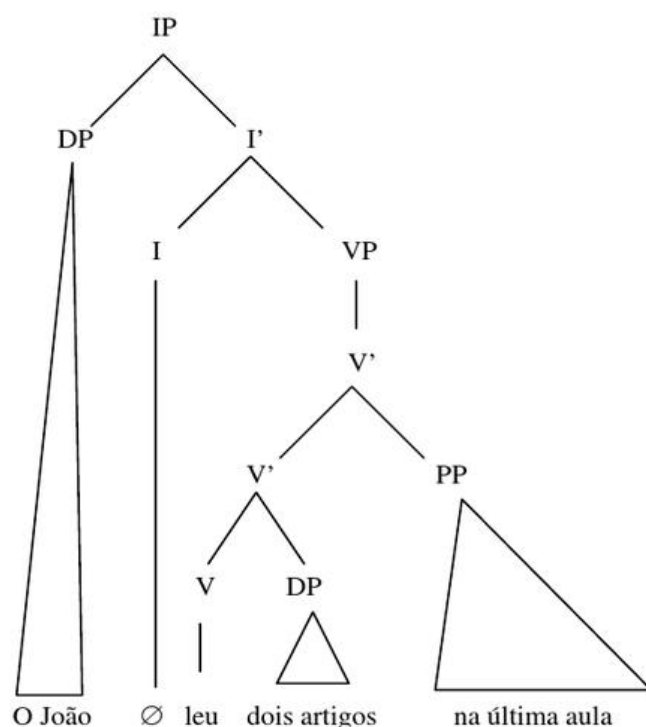
(a) I want **Mary** to drive me home.

(b) **Who** do you want $\mathbf{t_i}$ to drive you home?

(c) ***Who** do you wanna drive you home?

Já vimos desde o capítulo anterior que as análises que estamos desenvolvendo são análises da estrutura de superfície da sentença do PB, análises de **sintaxe mais simples**, por assim dizer. Não nos propusemos a analisar um outro nível de estrutura sintática que não aquele pronunciado e, mesmo assim – mesmo ficando com a análise do nível sintático de superfície –, não analisamos sentenças que passaram por processos de movimento e de reestruturação drásticos na ordem dos constituintes[123]. Estamos analisando a sentença do PB em sua ordem canônica e apenas em sua estrutura de superfície. Por isso, seguindo as análises da seção anterior, iremos considerar que o verbo (com sua flexão) ocupe a posição de núcleo do VP. A posição de núcleo do IP é ocupada por um verbo auxiliar, como já vimos, ou por um elemento vazio. O elemento vazio no núcleo do IP deve ser entendido como um morfema de flexão verbal, que, em nossas árvores, já aparecerá aglutinado ao verbo principal. Dessa forma, continuaremos a mostrar a análise sintática das sentenças em sua forma de superfície, em sua forma pronunciável.

Vejamos, então, a análise de uma sentença simples em PB:

(4.257) O João leu dois artigos na última aula.

[123] Não analisamos, por exemplo, sentenças que apresentam inversão predicado-sujeito (*Chegou a pizza*), estrutura de topicalização (*A Maria, eu vi* (*ela*) *no cinema ontem*), movimento-QU (*Que filme a Maria foi ver no cinema?*), etc. Analisamos apenas sentenças de superfície que passaram por operações de movimento que não tiveram grande influência na ordem dos elementos da sentença: alguns deslocamentos de advérbios, movimento do DP em posição de [Spec,VP] para [Spec,IP], etc.

Para analisar as sentenças simples, não precisamos elaborar nenhuma regra além daquelas que já havíamos apresentado. Para analisar sentenças simples com a ordem sujeito – predicado, precisamos apenas das regras já estudadas (4.246) (IP → DP I') e (4.248) (I' → I VP).

Vejamos a seguir a análise de sentenças que apresentam uma sequência do tipo **verbo pleno + verbo pleno**.

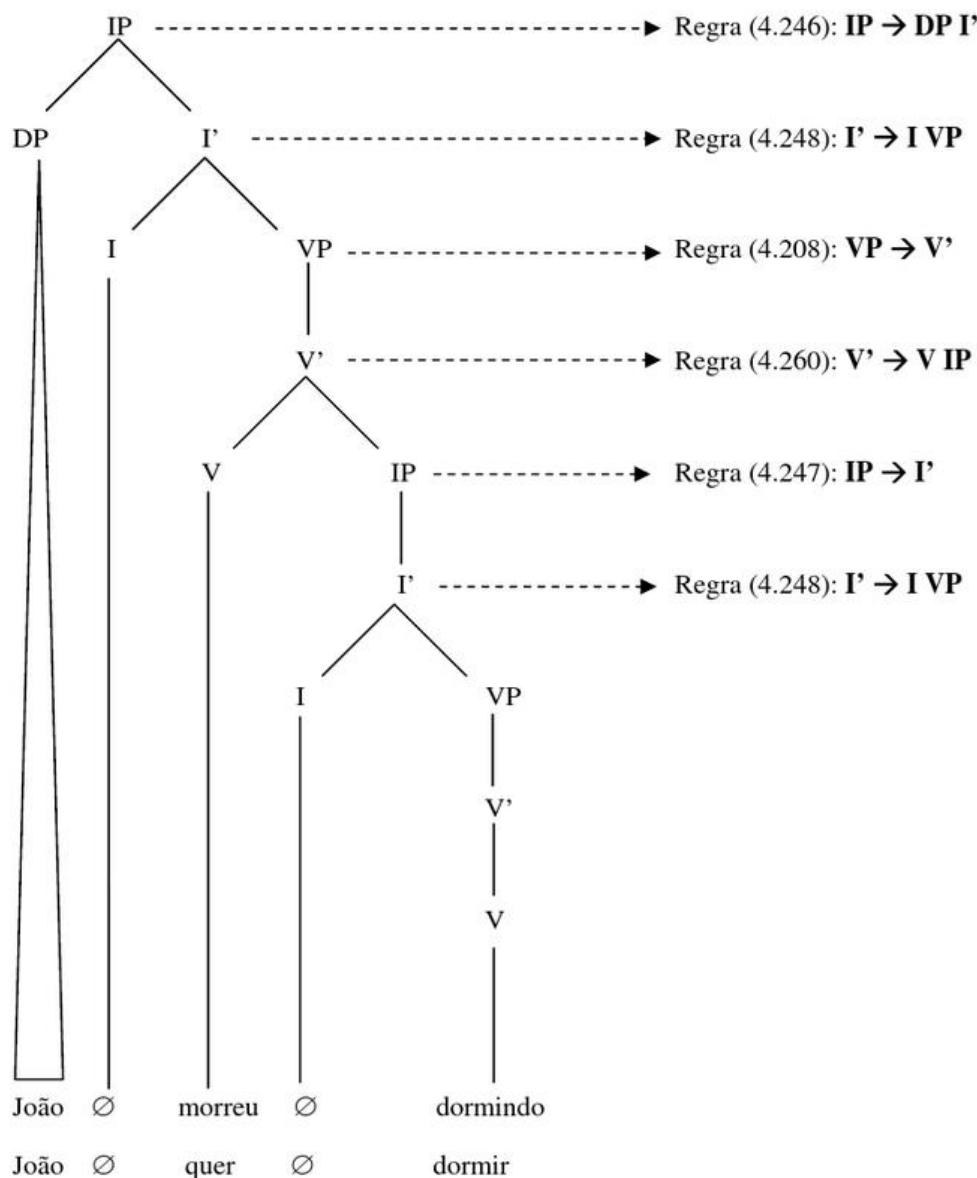
4.6.3 Sentenças com dois VPs

Com base nos testes que mostramos na seção 4.6.1, verificamos que, além das "locuções verbais", formadas por verbo auxiliar + verbo pleno, existem também sequências verbais que são compostas por dois verbos plenos, como vemos nos exemplos abaixo:

(4.258) João quer dormir.
(4.259) João morreu dormindo.

Para analisar essas estruturas, consideraremos cada verbo lexical como núcleo de seu sintagma verbal. Além das regras gramaticais que já apresentamos, a gramática deverá conter também a seguinte:

(4.260) V' → V IP

Essa regra formaliza a noção de que um IP pode servir como complemento verbal. Vejamos a análise das sentenças acima:

As duas sentenças que apresentamos acima servem de exemplo para o tipo de construção verbal que pode ser confundido com uma sequência do tipo verbo auxiliar + verbo principal. Da mesma maneira, essa mesma regra também pode analisar corretamente sentenças que apresentam um IP, flexionado ou não, com sujeito expresso em posição de complemento verbal, como mostraremos a seguir.

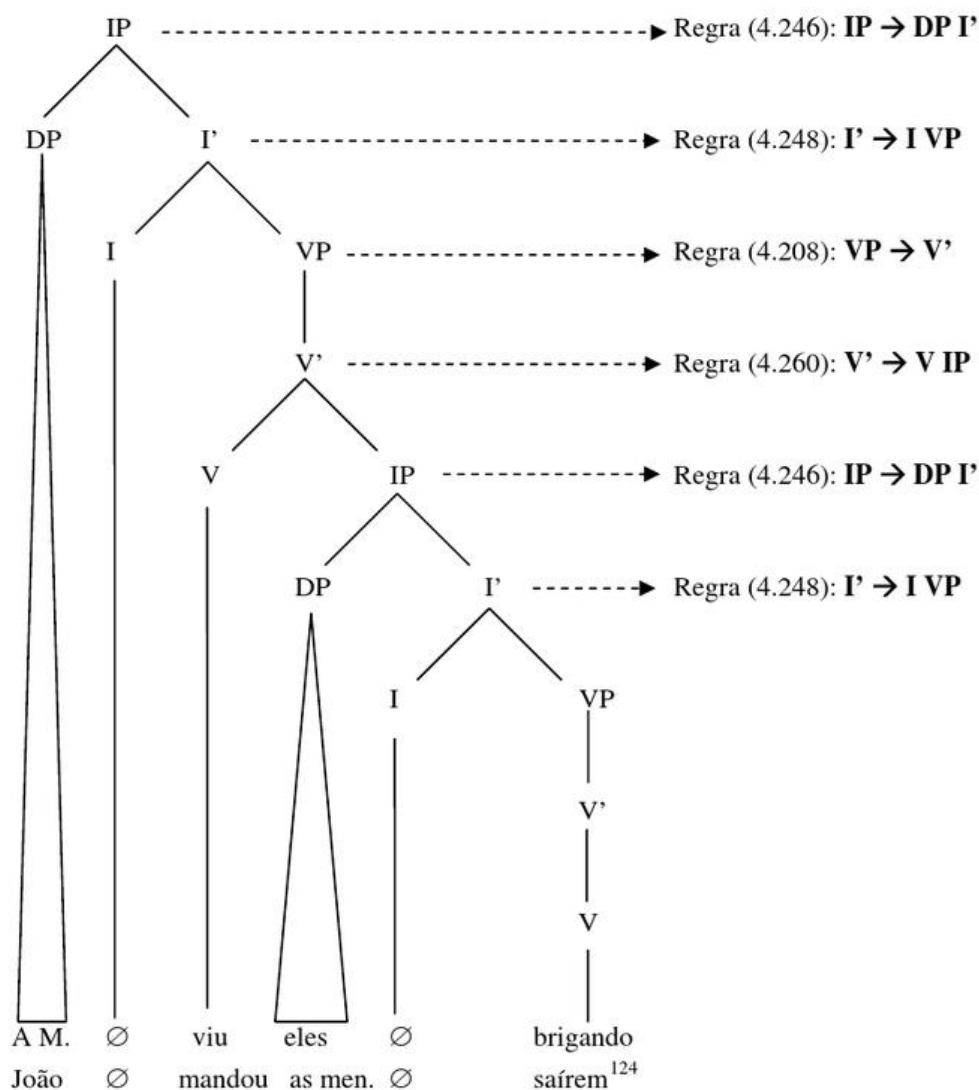
4.6.4 IPs e CPs como complementos do verbo

Na seção anterior, vimos que um IP pode ser complemento de verbo em uma

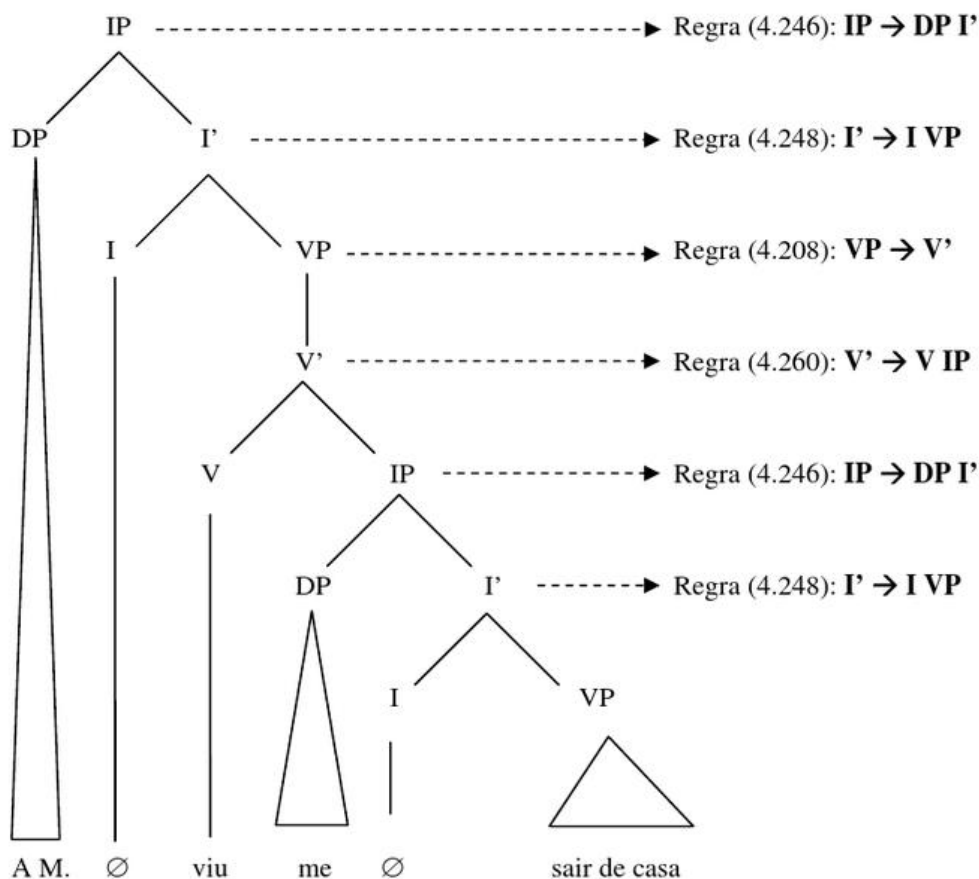
estrutura complexa quando temos dois verbos plenos no predicado da frase. Em português, pode acontecer também que o complemento do verbo seja uma oração (um IP) com sujeito em posição de especificador, [Spec, IP]. Isso pode acontecer com IPs gerundivos, como (4.261), ou IPs infinitivos, como (4.262).

(4.261) A Maria viu eles brigando.

(4.262) João mandou as meninas saírem.

As regras gramaticais que apresentamos até agora dão conta da análise desses tipos de sentença em PB. Confira a árvore sintática a seguir:

[124] Há autores, como Mioto et al. (2004), que sugerem que o complemento do V em frases como (4.262) seja um CP nulo (ou seja, um CP com o núcleo foneticamente vazio) que contém o IP. Não adotamos essa análise

Da mesma forma, podemos analisar sentenças com os pronomes em caso morfológico acusativo e com o verbo infinitivo não flexionado, como mostramos abaixo:

(4.263) A Maria viu-me sair de casa.

Em resumo: os complementos verbais oracionais podem estar em forma de orações infinitivas (flexionadas ou não) ou orações gerundivas. Ambos os tipos de complementos foram representados como IPs em posição de complementos verbais. Além de o complemento do verbo poder ser um IP, também as orações desenvolvidas, ou seja, também os CPs, podem ocupar essa posição. Vejamos os exemplos seguintes:

(4.264) João quer que a Maria chegue cedo.

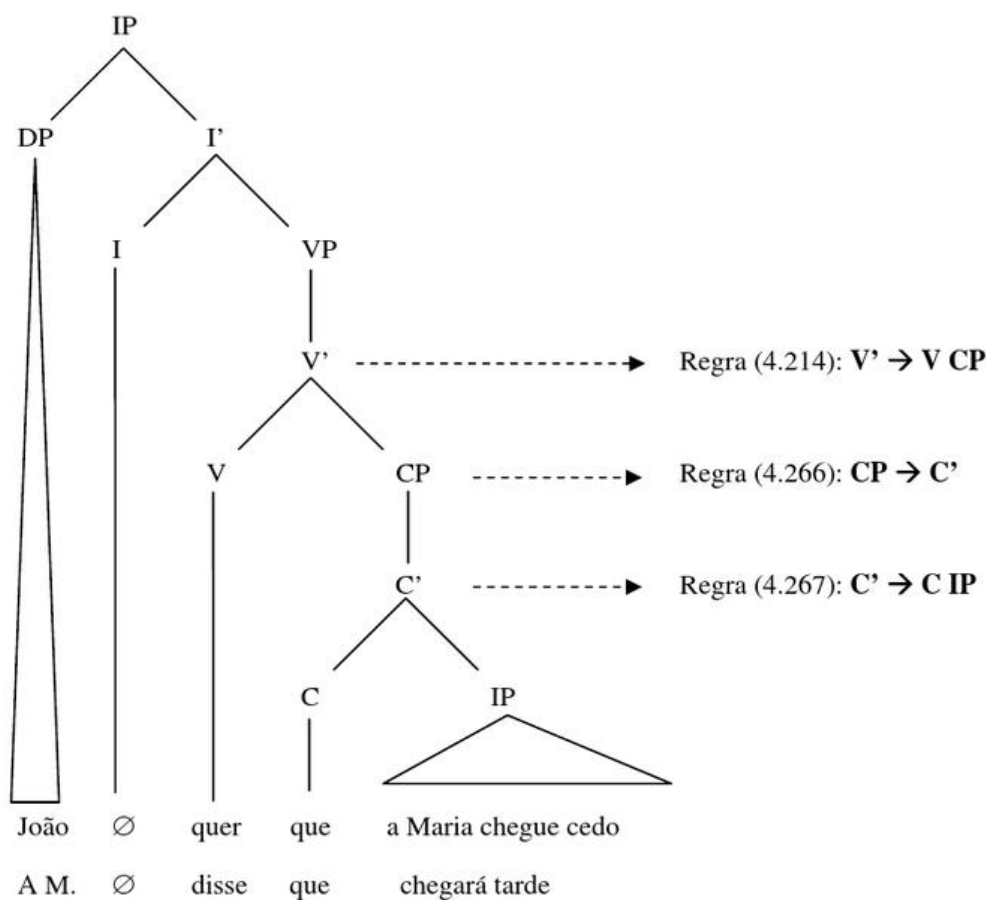
(4.265) A Maria disse que chegará tarde.

---

aqui por ser irrelevante à nossa discussão e por dificultar sem necessidade a análise sintagmática da sentença (cf. Mioto et al., 2004: 183-192).

Para analisar esse tipo de sentença, elaboramos as duas últimas regras de nossa gramática:

(4.266) CP → C'
(4.267) C' → C IP

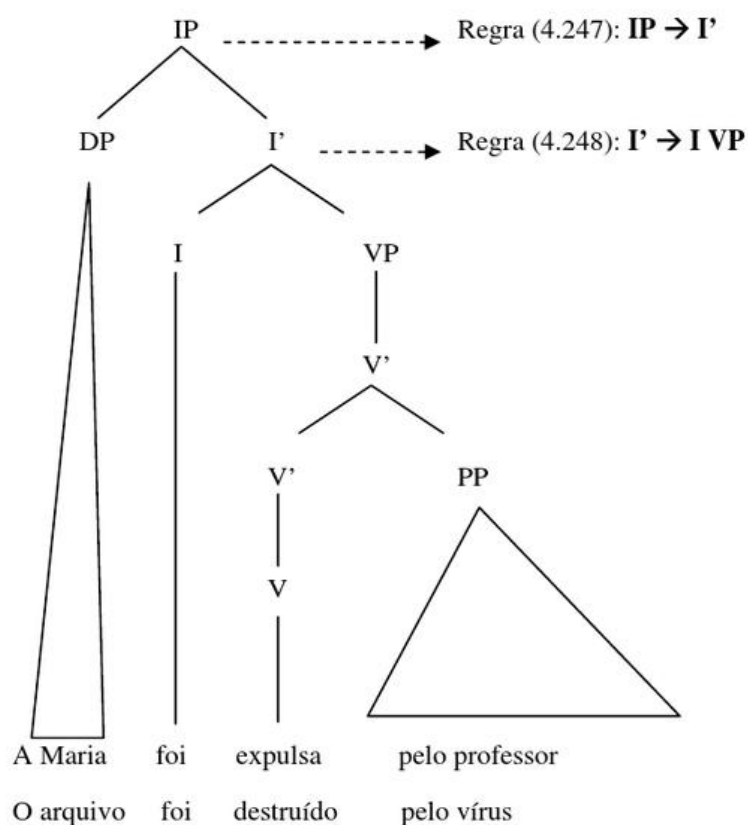
Vejamos a análise das duas frases anteriores:

A posição de núcleo do CP será ocupada por **que**, **se** ou **quando** (como em *João sabe **se** a Maria lê bastante*; *João sabe **quando** a Maria chegou*). Com essas mesmas regras, podemos analisar também frases que contenham uma estrutura verbo auxiliar + verbo principal dentro do IP complemento (*A Maria disse que iria chegar mais tarde*) e frases com CP recursivos (*A Maria disse que o João falou que o professor avisou que nós deveríamos estudar mais*).

4.6.5 Passivas

Ainda que seja considerada uma estrutura que tenha passado por operações de movimento[125], nossas regras gramaticais conseguem analisar corretamente as sentenças com estrutura passiva. Isso porque sua estrutura de superfície resultante é muito semelhante à estrutura das sentenças com verbos auxiliares, que analisamos no começo deste capítulo.

Vejamos dois exemplos:

(4.268) A Maria foi expulsa pelo professor.

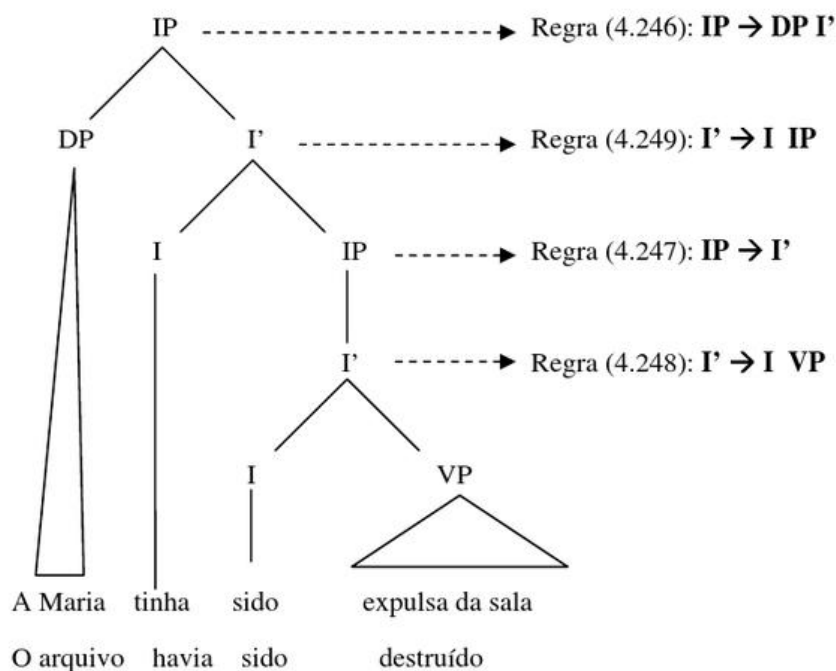
(4.269) O arquivo foi destruído pelo vírus.

O agente da passiva é analisado como adjunto em VP. Ele não é elemento obrigatório na frase, nem está presente no *grid* temático do verbo passivo[126]. As sentenças

[125] Cf. Baker (1988), Baker, Johnson & Roberts (1989).

[126] Cf. Haegeman (1995) e Carnie (2002), por exemplo.

passivas podem também apresentar mais de um verbo auxiliar, como vemos abaixo:

(4.270) A Maria tinha sido expulsa da sala.
(4.271) O arquivo havia sido destruído.

Para analisar essas frases, já temos as regras gramaticais de que precisamos. Sua análise será muito semelhante à análise das sentenças ativas com mais de um verbo auxiliar:

Na seção final deste capítulo, apresentamos todas as regras gramaticais que elaboramos. Acreditamos que essas regras possam descrever adequadamente as estruturas em PB que nos propusemos a estudar: os sintagmas lexicais e funcionais do PB e a sentença em sua ordem canônica (SVO), tanto a sentença simples (com apenas um verbo), como as sentenças com locuções verbais e com predicados complexos.

4.7 RESUMO DA GRAMÁTICA

Abaixo, listamos as 59 regras de descrição gramatical que apresentamos neste

capítulo.

DP → D'
D' → D
D' → D NP
D' → D PossP
D' → D NumP

NP → N'
N' → N
N' → N' AP
N' → AP N'
N' → N' PP
N' → N PP
N' → N' CP
N' → N CP

NumP → Num'
Num' → Num NP
Num' → Num PP

PossP → Poss'
Poss' → Poss NP
Poss' → NP Poss
Poss' → Poss NumP

QP → Q'
Q' → Q DP
Q' → Q PP
Q' → DP Q

Quadro 4.2: Regras gramaticais dos agrupamentos nominais

AP → A'
A' → A
A' → AdvP A'
A' → A' AdvP
A' → A' PP
A' → A PP
A' → A CP

Quadro 4.3: Regras gramaticais dos sintagmas adjetivais

PP → P'
P' → AdvP P'
P' → P DP
P' → P AdvP
P' → P CP
P' → P PP
P' → P

Quadro 4.4: Regras gramaticais dos sintagmas preposicionais

AdvP → Adv'
Adv' → AdvP Adv'
Adv' → Adv
Adv' → Adv PP

Quadro 4.5: Regras gramaticais dos sintagmas adverbiais

VP → V'
V' → AdvP V'
V' → V' AdvP
V' → V
V' → V DP
V' → V PP
V' → V CP
V' → V AP
V' → V AdvP
V' → V' PP
V' → V IP

Quadro 4.6: Regras gramaticais dos sintagmas verbais

IP → DP I'
IP → I'
I' → I VP
I' → I IP
CP → C'
C' → C IP

Quadro 4.7: Regras gramaticais da sentença

CONSIDERAÇÕES FINAIS

Uma das queixas que os linguistas por vezes recebem dos informatas interessados em trabalhar com a gramática da língua portuguesa é que faltam estudos gramaticais do português que apresentem descrições formalizadas passíveis de implementação computacional[127]. Ao longo de nosso estudo, propusemo-nos justamente a estudar alguns trabalhos de autores que já se preocuparam com a descrição da sintaxe do português brasileiro, tais como Pontes (1973), Lobato (1975, 1986), Luft (1986), Lemle (1984), Perini (1989, 2000) e Mioto et al. (2004). E partindo desses estudos, analisamos minuciosamente a estrutura dos sintagmas do PB. Acabamos propondo uma "gramática" do PB com 59 regras gramaticais, todas seguindo os moldes de uma teoria amplamente divulgada em Linguística, a teoria X-barra.

Nossas regras gramaticais são regras sintagmáticas de reescrita que, como vimos no capítulo 2, são passíveis de implementação computacional. Além de nosso estudo de Othero (2006), há outros estudos sobre *parsing* do português que já se beneficiaram de gramáticas com regras sintagmáticas de reescrita, tais como Pagani (2004), David (2007) e Othero & David (no prelo).

Analisamos com detalhe os sintagmas lexicais em português (NP, AP, AdvP, PP e VP) e utilizamos (ou propusemos) alguns sintagmas funcionais que auxiliaram nossas descrições (DP, NumP, PossP, QP). Além de estudar a estrutura interna desses constituintes, analisamos também a sentença em sua ordem canônica, SVO, e propusemos regras de descrição gramatical para a sentença em PB. Estudamos tanto as sentenças simples (aquelas com apenas um verbo), como sentenças complexas (com estruturas de verbo auxiliar + verbo principal e estruturas com predicados complexos).

Sabemos que o estudo gramatical que esboçamos aqui apresenta diversas limitações (muitas das quais já abordamos ao longo do texto). Mas esperamos ter avançado na compreensão do estudo da estrutura sintática da língua portuguesa, de maneira geral, e esperamos ter deixado aqui uma contribuição aos informatas que desejam se debruçar sobre a descrição sintática da língua portuguesa, especialmente em sua

[127] De fato, recentemente, durante o VIII Encontro do Círculo de Estudos Lingüísticos do Sul, realizado em outubro de 2008, em Porto Alegre, uma colega da Linguística Computacional da USP nos fez a seguinte pergunta: "que gramática da língua portuguesa que você recomenda para trabalhar com minhas aplicações computacionais?". Ela se queixava de haver poucos estudos gramaticais "formalizados" **e** "completos" **e** "acessíveis" para o português do Brasil – tarefa nada fácil de alcançar.

variedade brasileira.

Começamos nosso texto com uma citação de Perini. Felizmente – ou infelizmente –, podemos encerrar nosso trabalho com o mesmo trecho:

> Para quem gosta de certezas e seguranças, tenho más notícias: a gramática não está pronta. Para quem gosta de desafios, tenho boas notícias: a gramática não está pronta. Um mundo de questões e problemas continua sem solução, à espera de novas idéias, novas teorias, novas análises, novas cabeças. Perini (2003: 85)

REFERÊNCIAS BIBLIOGRÁFICAS

AARTS, Bas; HAEGEMAN, Liliane. English word classes and phrases. In: AARTS, Bas (ed.). *The handbook of English Linguistics*. Oxford: Blackwell, 2006.

ABNEY, Steven P. *The English noun phrase in its sentential aspect*. Department of Linguistics and Philosophy, Massachusetts Institute of Technology, 1987. Tese de Doutorado.

ALENCAR, Leonel F. Resenha de "Teoria X-barra: descrição do português e aplicação computacional". *ReVEL*, vol. 8, n. 10, 2008.

ALENCAR, Leonel F. Formalismos gramaticais baseados em estruturas de traços. In: ALENCAR, L. F.; OTHERO, G. A.; PAGANI, L. A. (orgs). *Abordagens computacionais da teoria da gramática*. (no prelo).

ARCHANGELI, Diana; LANGENDOEN, D. Terence (eds.) *Optimality Theory: an overview*. Oxford: Blackwell, 1997.

ARISTÓTELES. *Metafísica*. Porto Alegre: Globo, 1969 [350 a.C.].

ASSUNÇÃO, Carlos. As classes de palavras: dos primórdios a Cunha & Cintra. In: TORRES, M. A. E. et al. (eds.). *Estudios de Historiografía Lingüística. Actas del III Congreso Internacional de la Sociedad Española de Historiografia Lingüística*. Vigo, 7-10 de fevereiro, 2001.

AZEREDO, José Carlos. *Iniciação à sintaxe do português.* 6ª edição. Rio de Janeiro: Jorge Zahar, 2000.

AZEREDO, José Carlos. A palavra e suas classes. *Idioma* 21. Rio de Janeiro: Centro Filológico Clóvis Monteiro; Instituto de Letras – UERJ, 2001.

BAKER, Mark. *Incorporation: a theory of grammatical function in change*. Chicago: University of Chicago Press, 1988.

BAKER, Mark. *The atoms of language – the mind's hidden rules of grammar*. New York: Basic Books, 2001.

BAKER, M.; KYLE, J.; ROBERTS, I. Passive arguments raised. *Linguistic Inquiry* 20, 1989.

BATEMAN, J.; FORREST, J.; WILLIS, T. The use of syntactic annotation tools: partial and full parsing. In: GARSIDE, R.; LEECH, G.; McENERY, A. *Corpus annotation: linguistic information from computer text corpora*. London / New York: Longman, 1997.

BECHARA, Evanildo. *Lições de português pela análise sintática*. Rio de Janeiro: Fundo de Cultura, 1960.

BECHARA, Evanildo. *Nova gramática portuguesa*. São Paulo: Companhia Editora Nacional, 1963.

BECKER, Misha. *The development of the copula in child English: the lightness of be*. UCLA, 2000. Tese de Doutorado.

BERNSTEIN, Judy. DPs in French and Walloon: evidence for parametric variation in nominal head movement. *Probus* 3, 1991.

BICK, Eckhard. Automatic parsing of Portuguese. In: GARCÍA, L. S. (Ed.). *Anais do II Encontro para o Processamento Computacional de Português Escrito e Falado*. Curitiba: CEFET-PR. 1996.

BICK, Eckhard. *The parsing system Palavras - automatic grammatical analysis of Portuguese in a constraint grammar framework*. Aarhus: Aarhus University Press, 2000. Tese de Doutorado.

BLACK, Cheryl A. A PC-PATR implementation of GB syntax. *SIL Electronic Working Papers*, Dallas, v. 6, 1997.

BOFF, Raquel. *Em busca de uma análise sintático-semântica para construções com o verbo Começar no português brasileiro*. Paraná: Universidade Federal do Paraná, 2003. Dissertação de Mestrado.

BOLINGER, D.; SEARS, D. A. *Aspects of language*. New York: Harcourt Brace Jovanovich, 1968.

BORSLEY, Robert. *Syntactic theory: a unified approach*. London: Oxford University Press, 1999.

BRATKO, Ivan. *Prolog programming for artificial intelligence*. Harlow: Addison-Wesley, 1997.

BRITO, Ana Maria. European Portuguese possessives and the structure of DP. *Cuadernos de Lingüística* XIV, 2007.

BRNA, Paul. *Prolog programming: a first course*. [s.l.; s.ed.], 2001.

BURZIO, Luigi. *Italian syntax*. Dordrecht: Reidel, 1986.

CALLOU, Dinah et al. A posição do adjectivo no sintagma nominal: duas perspectivas de análise. In BRANDÃO, S. F.; MOTA, M. A. (orgs.). *Análise contrastiva de variedades do português - primeiros estudos*. Rio de Janeiro: In-Fólio, 2003.

CÂMARA JR., J. Mattoso. *Estrutura da língua portuguesa*. Petrópolis: Vozes, 1970.

CÂMARA JR., J. Mattoso. *Dicionário de Lingüística e gramática*. Petrópolis: Vozes, 1981.

CANÇADO, Márcia. Predicação, relações semânticas e papéis temáticos. Anotações de Carlos Franchi. *Revista de Estudos da Linguagem*. Vol. 11, n. 2, 2003.

CANÇADO, Márcia. Argumentos: complementos e adjuntos. *Oficina de Semântica NUPES/POSLIN/UFMG*, maio de 2005.

CARNIE, Andrew. *Syntax: a generative introduction*. Oxford: Blackwell, 2002.

CASTRO, Ana. *On possessives in Portuguese*. Universidade Nova de Lisboa/Université Paris 8, 2006a. Tese de Doutorado.

CASTRO, Ana. Possessivos e artigo definido expletivo em PE e PB. *Veredas*, vol. 10, n. 1 e 2, 2006b.

CEGALLA, Domingos Pascoal. *Novíssima gramática da língua portuguesa*. São Paulo: Nacional, 1996.

CHEN, Hsin-Hsi; LIN, I. Peng; WU, Chien-Ping. A new design of prolog-based bottom-

up parsing system with Government-Binding theory. *Proceedings of the 12th conference on Computational Linguistics*. Budapest, Hungary, 1988.

CHOMSKY, Noam. *The logical structure of linguistic theory*. MIT, 1955. Tese de Doutorado.

CHOMSKY, Noam. Three models for the description of language. *IRE Transactions on Information Theory*, vol. IT-2, 1956.

CHOMSKY, Noam. *Syntactic structures*. The Hague: Mouton 1957.

CHOMSKY, Noam. *Aspects of the theory of syntax*. Cambridge: MIT Press, 1965.

CHOMSKY, Noam. Remarks on nominalization. In: JACOBS, R.; ROSENBAUM, P. (eds.) *English Transformational Grammar*. Washington: Georgetown University Press, 1970.

CHOMSKY, Noam. *Lectures on Government and Binding*. Dordrecht: Foris, 1981.

CHOMSKY, Noam. *Knowledge of language: its nature, origin and use*. New York: Praeger, 1986a.

CHOMSKY, Noam. *Barriers*. Cambridge: MIT Press, 1986b.

CHOMSKY, Noam. *Language and problems of language – the Managua lectures*. Cambridge: MIT Press, 1988.

CHOMSKY, Noam. *The minimalist program*. Cambridge: MIT Press, 1995.

CHOMSKY, Noam. *On nature and language*. Cambridge: Cambridge University Press, 2002.

CHOMSKY, Noam; LASNIK, Howard. Filters and Control. *Linguistic Inquiry*, 8, 1977.

CINQUE, Guglielmo. On the evidence for partial N-movement in the Romance noun phrase. In: CINQUE, G. (ed.) *Italian syntax and Universal Grammar*. Cambridge: Cambridge University Press, 1995.

CINQUE, Guglielmo. *Adverbs and functional heads. A cross-linguistic perspective*. Oxford: Oxford University Press, 1999.

CINQUE, Guglielmo. Deriving Greenberg's universal 20 and its exceptions. *Linguistic Inquiry* 36, 2, 2005.

COENE, Martine; D'HULST, Yves (orgs.). *From NP to DP. Vol. 1: the syntax and semantics of noun phrases*. Amsterdam: Benjamins, 2003a.

COENE, Martine; D'HULST, Yves (orgs.). *From NP to DP. Vol. 2: the expression of possession in noun phrases*. Amsterdam: Benjamins, 2003b.

CORSO, Sofia. *Verbos auxiliares no português brasileiro*. Florianópolis: Universidade Federal de Santa Catarina, 2002. Dissertação de Mestrado.

COSTA, João. Positions for subjects in European Portuguese. *Proceedings of WCCFL XV*, CSLI, Stanford University, 1997.

COSTA, João. *Word order variation – a constraint-based approach*. The Hague: Holland Academic Graphics, 1998.

COSTA, Ana; COSTA, João. *O que é um advérbio?*. Lisboa: Edições Colibri, 2001.

CULICOVER, Peter W. *Principles and parameters – an introduction to syntactic theory*. Oxford: Oxford University Press, 1997.

CULICOVER, Peter W.; JACKENDOFF, Ray. *Simpler syntax*. Oxford/New York: Oxford University Press, 2005.

DAVID, Karine Alves. O pronome pessoal enquanto núcleo de um DP: uma abordagem computacional. *Semana de Letras da UFC*, 2006a.

DAVID, Karine Alves. Uma análise em Prolog do pronome pessoal enquanto núcleo do DP. *XXI Jornada Nacional de Estudos Lingüísticos*, 2006b.

DAVID, Karine Alves. *Sintaxe das expressões nominais no português do Brasil: uma abordagem computacional*. Programa de Pós-Graduação em Lingüística, Universidade Federal do Ceará, 2007. Dissertação de Mestrado.

DUTRA, Rosália. *O falante gramático: introdução à prática de estudo e ensino do português*. Campinas: Mercado de Letras, 2003.

EMMONDS, Joseph E. *A transformational approach to English syntax*. New York: Academic Press, 1976.

ERNST, Thomas. *The syntax of adjuncts*. Cambridge: Cambridge University Press, 2001.

ESCRIBANO, José L. G. NPs as just NPs. *Language Sciences*, 28, 2006.

FALK, Yehuda N. *Structure of "Sentence". Structure and Function in Syntax*. Handout. Department of English, The Hebrew University of Jerusalem, 2006.

FIGUEIREDO SILVA, Maria Cristina; ARAÚJO, Flávio Martins. Prosódia e ordem das palavras: o caso dos advérbios ditos monossilábicos átonos. *ReVEL*, vol. 6, n. 10, 2008.

FONG, Sandiway. Parallel Principle-Based Parsing. In: *Proceedings of Natural Language Understanding and Logic Programming (NLULP)*, International Conference on Logic Programming (ICLP), Las Cruces, New Mexico, 1999.

FONG, Sandiway. Japanese PAPPI. In: *Researching and verifying an advanced theory of human language*, Report (5), Kanda University of International Studies (KUIS), Chiba, Japan, 2001.

FOUSKAKIS, Konstantinos. A computational methodology as the artificial language for natural language rules and the unification based approach. *Second Romanian-Hungarian Joint Symposium on Applied Computational Intelligence*. Romania, 2005.

FUKUI, N.; SPEAS, M. Specifiers and projection. In: FUKUI, N.; RAPOPORT, T. R.; SAGEY, E. (orgs). *MIT Working Papers in Linguistics*, vol. 8, MIT, Cambridge, 1986.

GARSIDE, R.; LEECH, G.; McENERY, A. *Corpus annotation: linguistic information from computer text corpora*. London / New York: Longman, 1997.

GAZDAR, Gerald. Phrase Structure Grammars and natural language. *Proceedings of the Eighth International Joint Conference on Artificial Intelligence*, v.1, August, 1983.

GAZDAR, G.; PULLUM, G.; SAG, I. Auxiliaries and related phenomena in a restrictive theory of grammar. *Language* 58, 1982.

GAZDAR, G.; KLEIN, E.; PULLUM, G.; SAG, I. *Generalized Phrase Structure Grammar*. Oxford: Basil Blackwell, 1985.

GIORGI, Alessandra; LONGOBARDI, Giuseppe. *The syntax of noun phrases*. Cambridge: Cambridge University Press, 1991.

GIUSTI, Guiliana. The categorial status of determiners. In: HAEGEMAN, L. (ed.).*The new comparative syntax*. London: Longman, 1997.

GLANZBERG, Michael. Metaphor and lexical semantics. In: SKILTERS, J. et al. *The Baltic International Yearbook of Cognition, Logic and Communication*. University of Latvia, 2008.

GONÇALVES, Anabela. *Para uma sintaxe dos verbos auxiliares em português europeu*. Faculdade de Letras da Universidade de Lisboa, Lisboa, 1992. Dissertação de Mestrado.

GONÇALVES, Anabela. Aspectos da sintaxe dos verbos auxiliares em português europeu. In: Gonçalves, A. et al. *Quatro estudos em sintaxe do português*. Lisboa: Colibri, 1996.

GONZAGA, Manuela. The structure of DP in European Portuguese. Evidence from adjectives and possessives. *First Annual Graduate Student Colloquium at Harvard*, Harvard, 2003.

GONZAGA, Manuela. On the left periphery of DP – evidence from possessives. *LESCOL 2006*, Leipzig, 2006.

GREWENDORF, Günther; HAMM, Fritz; STERNEFELD, Wolfgang. *Sprachliches Wissen: Eine Einführung in moderne Theorien der grammatischen Beschreibung*. Frankfurt am Main: Suhrkamp, 1989.

GRISHAM, Ralph. *Computational linguistics: an introduction*. Cambridge: Cambridge University Press, 1992.

HAEGEMAN, Liliane. *Introduction to government and binding theory*. Oxford: Blackwell, 1995.

HARRIS, Randy Allen. *The linguistics wars*. New York: Oxford University Press, 1993.

HAWKINS, John A. *A performance theory of order and constituency*. Cambridge: Cambridge University Press, 1994.

HAWKINS, John A. Some issues in a performance theory of word order. In: SIEWIERSKA, A. *Constituent order in the languages of Europe*. Berlin: De Gruyter, 1997.

HODGES, Andrew. *Alan Turing: the enigma*. New York: Simon and Schuster, 1983.

HOPCROFT, J. E.; MOTWANI, R.; ULLMAN, J. D. *Introduction to automata theory, languages, and computation*. 2nd edition. Addison-Wesley, 2001.

HUDDLESTON, Rodney. *Introduction to the grammar of English*. Cambridge: Cambridge University Press, 1984.

HUTCHINS, W. J. (ed.). *Early years in machine translation* Amsterdam: John Benjamins, 2000.

ILARI, Rodolfo. A perspectiva funcional da frase portuguesa. Campinas: Ed. da Unicamp, 1992.

ILARI, Rodolfo (org.) *Gramática do português falado, vol. II: níveis de análise lingüística*. 4ª edição. Campinas: Editora da Unicamp, 2002.

ILARI, Rodolfo et al. Considerações sobre a posição dos advérbios. In: CASTILHO, A. T. (org.). *Gramática do português falado, vol. I: A ordem*. 4ª edição. Campinas: Editora da Unicamp, 1990.

JACKENDOFF, Ray. *Semantic interpretation in generative grammar*. Cambridge: MIT Press, 1972.

JACKENDOFF, Ray. The base rules for prepositional phrases. In: ANDERSON, S.; KIPARSKY, P. (orgs.). *A Festchrift for Morris Halle*. New York: Holt, Rinehart and Wilson, 1973.

JACKENDOFF, Ray. *X' syntax: a study of phrase structure*. Cambridge: MIT Press, 1977.

JACKENDOFF, Ray. *Foundations of language: brain, meaning, grammar, evolution*. Oxford: Oxford University Press, 2002.

KAC, Michael B.; MANASTER-RAMER, Alexis. Parsing without (much) phrase structure. *Proceedings of the 11th conference on Computational Linguistics*. Bonn, Germany, 1986.

KAYNE, Richard. *Connectedness and binary branching*. Dordrecht: Foris, 1984.

KAYNE, Richard. *The antisymmetry of syntax*. Cambridge: MIT Press, 1994.

KLEIN, Susan. Foco no português brasileiro. In: MÜLLER, A. L.; NEGRÃO, E.V.; FOLTRAN, M. J. *Semântica Formal*. São Paulo: Contexto, 2003.

KORNAI, András; PULLUM, Geoffrey K. The X-bar theory of phrase structure. *Language* 66, 1990.

KUHN, Jonas. *Optimality-theoretic syntax - a declarative approach*. CSLI Publications, 2003.

KUHNS, Robert J. A PROLOG implementation of government-binding theory. *Proceedings of the 11th conference on Computational Linguistics*. Bonn, Germany, 1986.

KUHNS, Robert J. Automatic indexing and government-binding theory. *Proceedings of the 13th conference on Computational Linguistics*. Helsinki, Finland, 1990.

LAKOFF, George. Global rules. *Language*, 46, 1970.

LARSON, Richard. K. On the double object construction. *Linguistic Inquiry*, 19, 1988.

LARSON, Richard. K. Double objects revisited: reply to Jackendoff. *Linguistic Inquiry* 21, 1990.

LEGENDRE, Géraldine. An introduction to optimality theory in syntax. In: LEGENDRE, G.; GRIMSHAW, J.; VIKNER, S. (eds.). *Optimality-Theoretic Syntax*. MIT Press, 2001.
LEMLE, Miriam. *Análise sintática: teoria geral e descrição do português*. São Paulo: Ática, 1984.

LEVIN, Beth. *English verb classes and alternations*. Chicago: University of Chicago Press, 1993.

LOBATO, Lúcia M. P. Os verbos auxiliares em português contemporâneo. Critério de auxiliaridade. In: LOBATO, L. M. P. et al. *Análises lingüísticas*. Petrópolis: Vozes, 1975.

LOBATO, Lúcia M. P. *Sintaxe gerativa do português: da Teoria Padrão à Teoria da Regência e Ligação*. Belo Horizonte: Vigília, 1986.

LUFT, Celso Pedro. *Moderna gramática brasileira*. 7ª edição. Porto Alegre; Rio de Janeiro: Globo, 1986.

LUNGUINHO, Marcus Vinícius. A sintaxe diacrônica de ter de/ter que. *IV Encontro Nacional do Grupo de Estudos de Linguagem do Centro-Oeste*, 2008a.

LUNGUINHO, Marcus Vinícius. A sintaxe dos verbos auxiliares de complementação não-imediata. *VII Encontro do Círculo de Estudos Lingüísticos do Sul – CELSUL*, 2008b.

LUNGUINHO, Marcus Vinícius. *Verbos auxiliares e a teoria da gramática: morfossintaxe, aquisição e diacronia*. São Paulo: Universidade de São Paulo, (a sair). Tese de Doutorado.

MARTINHO, Fernando Jorge dos Santos. *A elipse nominal em português e francês*. Faculdade de Letras da Universidade do Porto, 1998. Dissertação de Mestrado.

MATEUS, Maria Helena Mira; XAVIER, Maria Francisca (Orgs). *Dicionário de termos lingüísticos*. Lisboa: Edições Cosmos, 1992.

MATTOS E SILVA, Rosa V. *O português arcaico: fonologia, morfologia e sintaxe*. São Paulo: Contexto, 2006.

McDONALD, Carlton; YAZDANI, Masoud. *Prolog programming: a tutorial introduction*. Oxford: Blackwell Scientific Publications, 1990.

MENUZZI, Sérgio de Moura. *Sobre a modificação adjetival em português*. Campinas: Unicamp, 1992. Dissertação de Mestrado.

MENUZZI, Sérgio de Moura. *Sobre a perspectiva funcional da frase em português: revisitando o papel das "estruturas sintáticas marcadas"*. Projeto de produtividade em pesquisa submetido ao CNPq. Porto Alegre: Programa de PG em Letras da PUCRS, 2004.

MENUZZI, Sérgio; MIOTO, Carlos. Advérbios monossilábicos pós-verbais no PB: sobre a relação entre sintaxe e prosódia. *Revista de Estudos da Linguagem*, v. 14, n. 2, 2006.

MENUZZI, S. M.; OTHERO, G. A. Sintaxe X-barra: uma aplicação computacional. *Working papers em Lingüística*. UFSC, 2008.

MIOTO, C.; SILVA, M. C. F.; LOPES, R. E. V. *Novo manual de sintaxe*. Florianópolis: Insular, 2004.

MOLSING, Karina. Sobre a aquisição L2 de uma língua semi pro-drop. *I Colóquio de Lingüística & Literatura PUCRS*. Porto Alegre, PUCRS, Faculdade de Letras, 2008.

MONTALBETTI, Mario. *After binding: on the interpretation of pronouns*. Massachusetts Institute of Technology, 1984. Tese de Doutorado.

MORAES DE CASTILHO, Célia Maria. Primeiras histórias sobre a diacronia do dequeísmo: o clítico locativo *en* e o dequeísmo em orações relativas no PM. In: LOBO, T. et al. (orgs.). *Para a história do português brasileiro: novos dados, novas análises*. Salvador: EDUFBA, 2006.

MOURA NEVES, Maria Helena. Possessivos. In: CASTILHO, Ataliba Teixeira (org.). *Gramática do português falado, vol. III: as abordagens*. Campinas: Editora da Unicamp, 1993.

MOURA NEVES, Maria Helena. *Gramática de usos do português*. São Paulo: UNESP, 2000.

NEWMEYER, Frederick. J. *Linguistic theory in America: the first quarter-century of transformational generative grammar*. New York: Academic Press, 1980.

NEWMEYER, Frederick. J. *Generative grammar and functional explanation*. Handout. (s.l) (s.d.).

NIJHOLT, A.; SIKKEL, K. *Natural language processing: methods and formalisms. Proceedings Twente workshop on language technology 6*. University of Twente, 1993.

OPPY, Graham; DOWE, David. The Turing Test. In: ZALTA, E. (org.) *The Stanford encyclopedia of philosophy*. The Metaphysics Research Lab, Stanford, 2005.

OTHERO, Gabriel de Ávila. *Grammar Play: um* parser *sintático em Prolog para a língua portuguesa*. Porto Alegre: PUCRS, 2004. Dissertação de Mestrado.

OTHERO, Gabriel de Ávila. *Teoria X-barra: descrição do português e aplicação computacional*. São Paulo: Contexto, 2006.

OTHERO, Gabriel de Ávila. Lingüista "puro" vs. lingüista "computacional": revisitando a distinção entre "lingüista de poltrona" e "lingüista aplicado". *Domínios de Lingu@gem – Revista Eletrônica de Lingüística*. Ano 2, nº 1, 2008.

OTHERO, G. A.; DAVID, K. A. Agrupamentos nominais no português brasileiro: uma análise formal e sua aplicação computacional. In: ALENCAR, L. F.; OTHERO, G. A.; PAGANI, L. A. (orgs). *Abordagens computacionais da teoria da gramática*. (no prelo).

PAGANI, Luiz Arthur. Analisador gramatical em Prolog para gramáticas de estrutura sintagmática. *ReVEL*, vol. 2, n. 3, 2004.

PEREIRA, Eduardo Carlos. *Gramática expositiva*. São Paulo: Editora Nacional, 1958 [1918].

PEREIRA, F.; SHIEBER, S. M. *Prolog and natural-language analysis*. Stanford: CSLI, 1987.

PEREIRA, F.; WARREN, D. Definite clause grammars for language analysis: a survey of the formalism and a comparison with augmented transition networks. *Artificial Intelligence*, n. 13, 1980.

PERELTSVAIG, Asya. *On the universality of DP: a view from Russian*. Manuscript, Cornell University, 2005.

PERINI, Mário A. *Sintaxe portuguesa: metodologia e funções*. São Paulo: Ática, 1989.

PERINI, Mário A. *Gramática descritiva do português*. 4ª edição. São Paulo: Ática, 2000.

PERINI, Mário A. *Sofrendo a gramática*. 3ª edição. São Paulo: Ática, 2003.

PERINI, Mário A. *Princípios de lingüística descritiva: introdução ao pensamento gramatical*. São Paulo: Parábola, 2006.

PESETSKY, David. *Paths and categories*. MIT, 1982. Tese de Doutorado.

PEZZATI, E. G.; CAMACHO, R. G. Aspectos funcionais da ordem de constituintes. *DELTA*, vol. 13 n. 2, 1997.

POLLARD, C.; SAG, I. *Information-based syntax and semantics*. Vol. 1. Stanford: CSLI, 1987.

POLLARD, C.; SAG, I. *Head-driven Phrase Structure Grammar*. Chicago: University of Chicago Press, 1994.

PONTES, Eunice. *Verbos auxiliares no português*. Petrópolis: Vozes, 1973.

PULLUM, Geoffrey K.; GAZDAR, Gerald. Natural languages and context-free languages. *Linguistics and Philosophy* 4, 1982.

QUIRK, R.; GREENBAUM, S.; LEECH, G.; SVARTVIK, J. *A comprehensive grammar of the English language*. London: Longman, 1985.

RADFORD, Andrew. *Transformational syntax: a student's guide to Chomsky's extended standard theory*. Cambridge: Cambridge University Press, 1981.

RADFORD, Andrew. *Transformational grammar*. Cambridge: Cambridge University Press, 1988.

RADFORD, Andrew. *Syntactic theory and the structure of English: a minimalist approach*. Cambridge: Cambridge University Press, 1997a.

RADFORD, Andrew. *Syntax: a minimalist introduction*. Cambridge: Cambridge University Press, 1997b.

RAPOSO, Eduardo. P. *Teoria da gramática. A faculdade da linguagem*. Lisboa: Caminho, 1992.

ROCHA LIMA, Carlos Henrique. *Gramática normativa da língua portuguesa*. 24a ed. Rio Janeiro: José Olympio, 1984.

RODRIGUES, Rosany Schwarz. *Testando a distinção tradicional entre complementos e adjuntos nominais*. Porto Alegre: PUCRS, 2000. Dissertação de Mestrado.

ROTHMAN, Jason; IVERSON, Michael. On the syntax of null subjects in L2 Spanish: comparing two L2 populations under different exposure. *RESLA, Spanish Journal of Applied Linguistics*, vol. 20, 2007.

RUSSEL, S.; NORVIG, P. *Artificial intelligence*. New Jersey: Prentice-Hall, 1995.

SAG, I.; WASOW, T.; BENDER, E. *Syntactic theory: a formal introduction*. Stanford: CSLI Publications, 2003.

SANTORINI, Beatrice; KROCH, Anthony. *The syntax of natural language: an online introduction using the Trees program*. University of Pennsylvania, 2007.

SARAIVA, Fernanda C. Advérbios, uma abordagem crítica. *Soletras*. Ano VIII, n. 15. São Gonçalo: UERJ, 2008.

SAVITCH, W.; BACH, E.; MARSH, W.; SAFRAN-NAVEH, G. *The formal complexity of natural language*. Dordrecht: D. Reidel, 1987.

SHIEBER, Stuart M. Evidence against the context-freeness of natural language. *Linguistics and Philosophy* 8, 1985.

SHLONSKY, U. Quantifiers as functional heads: a study of quantifier float in Hebrew. *Lingua* 84, 1991.

SOAMES, Scott; PERLMUTTER, David M. *Syntactic argumentation and the structure of English*. Berkeley: University of California Press, 1979.

SOUZA e SILVA, Cecília P. de; KOCH, Ingedore Villaça. *Lingüística aplicada ao português: sintaxe*. São Paulo: Cortez, 1993.

SPARCK JONES, Karen. Natural Language Processing: a historical review. In ZAMPOLI, A.; CALZOLARI, N.; PALMER, M. (eds.). *Current issues in Computational Linguistics: in honour of Don Walker*. Amsterdam: Kluwer, 1994.

STOWELL, Timothy. *Origins of phrase structure*. Massachusetts: MIT, 1981. Tese de Doutorado.

STROZER, J. R. *Language acquisition after puberty*. Washington, DC: Georgetown University Press, 1994.

STUURMAN, F. *Phrase structure theory in generative grammar*. Dordrecht: Foris, 1985.

THOMAS, Margaret. *Universal Grammar in second language acquisition: a history*. New York: Routledge, 2004.

TURING, Alan. Computing machinery and intelligence. *Mind*, 59, 1950.

VALOIS, Daniel. The internal syntax of DP and adjective placement in French and English. *The Graduate Students Linguistic Association*, 1991.

VANGSNES, Øystein Alexander. On noun phrase architecture, referentiality, and article systems. *Studia Linguistica*, 55(3), 2001.

VICENTE, Helena da Silva Guerra. *O quantificador flutuante todos no português brasileiro e no inglês: uma abordagem gerativa*. Universidade de Brasília, 2006. Tese de Doutorado.

WILLIAMS, Edwin. Small clauses in English. In KIMBALL, J. (ed.), *Syntax and semantics*. New York: Seminar Press, 1975.

WINOGRAD, Terry. *Language as a cognitive process. Vol. I: Syntax*. Massachussetts: Addison-Wesley, 1983.

WONG, Aboy; WU, Dekai. Are Phrase Structured Grammars useful in statistical parsing?. *Proceeding of 5th Natural Language Processing Pacific Rim Symposium*, 1999.